

NOISE ESTIMATION IN THE SPIKED COVARIANCE MODEL

DIDIER CHÉTELAT AND MARTIN T. WELLS

Department of Statistical Science, Cornell University

ABSTRACT. The problem of estimating a spiked covariance matrix in high dimensions under Frobenius loss, and the parallel problem of estimating the noise in spiked PCA is investigated. We propose an estimator of the noise parameter by minimizing an unbiased estimator of the invariant Frobenius risk using calculus of variations. The resulting estimator is shown, using random matrix theory, to be strongly consistent and essentially asymptotically normal and minimax for the noise estimation problem. We apply the construction to construct a robust spiked covariance matrix estimator with consistent eigenvalues.

1. INTRODUCTION

The estimation of covariance matrices in a high dimensional framework has seen a surge of interest in the past years. The natural estimator, the sample covariance matrix, is well known to be inadequate in this context. The problem has been well studied under many sparsity scenarios: for example, zeros in the coordinates of the matrix [Bickel and Levina, 2008b, El Karoui, 2008b, Rothman et al., 2009, Cai and Liu, 2011] or its inverse [Meinshausen and Bühlmann, 2006, Friedman et al., 2008, Cai et al., 2011, Ravikumar et al., 2011, Rothman et al., 2008], bandedness [Bickel and Levina, 2008a, Bien et al., 2014] and many others. This paper will focus on the spiked model, first introduced by Johnstone [2001].

In the spiked model, the $p \times p$ covariance matrix Σ has distinct eigenvalues $\gamma_1 + \sigma^2 > \dots > \gamma_\rho + \sigma^2$, and a smallest eigenvalue σ^2 of multiplicity $p - \rho$. It often provides good approximations in low and high dimensional settings, with small ρ being seen as a form of low rank sparsity in the data. It is also of substantial theoretical interest, being one of the few non-trivial settings in which random matrix theory has been extensively studied.

E-mail address: dc623@cornell.edu, mtw1@cornell.edu.

2000 Mathematics Subject Classification. Primary 62F10; secondary 62H25, 62H12.

Key words and phrases. Covariance estimation, high-dimensional asymptotics, principal components, random matrix theory, risk function, Wishart distribution, spiked covariance.

This research was partially supported by NSF Grants DMS-1208488 and CCF-0808864.

A related problem is principal components analysis. In PCA, one estimates eigenvectors associated with large eigenvalues of Σ , and perform dimension reduction using a truncated spectral decomposition. A traditional problem with the technique is that the number of eigenvectors to retain is not clear. However, if the true covariance matrix Σ is spiked, it is natural to associate its spiked rank ρ with the ideal number of eigenvectors to select, recasting the selection of the number of components as a rigorous statistical estimation problem.

Successful high-dimensional PCA usually requires good estimation of σ^2 (see e.g. [Johnstone and Lu \[2009\]](#)), a problem we will refer to as noise estimation. Although distinct from estimation of the covariance matrix itself, there is a context in which these two problems, estimation of Σ and σ^2 , are analogous.

This context is as follows. Asymptotics are high-dimensional in the sense that p tends to infinity with the sample size n ; for mathematical convenience we focus on the regime where the ratio p/n tends to a strictly positive constant as $n \rightarrow \infty$. The noise estimation problem is to estimate σ^2 under, say, absolute error loss $L(\hat{\sigma}^2, \sigma^2) = |\hat{\sigma}^2 - \sigma^2|$, while the covariance problem is to estimate the spiked Σ under the Frobenius loss $L_F(\hat{\Sigma}, \Sigma) = \|\hat{\Sigma} - \Sigma\|_F^2/p$ using a spiked estimator. This normalization is natural in this setting, since under normality the risk $E[L_F(S, \Sigma)]$ of the sample covariance matrix S tends to a strictly positive constant.

Then, in essence, all that really matters in the covariance estimation problem is estimation of the noise level. Indeed, consider two spiked estimators $\hat{\Sigma}_i = \hat{\Gamma}_i + \hat{\sigma}^2 I$, $i = 1, 2$ with asymptotically finite spiked parts $\hat{\Gamma}_i$, which we take to mean that their ranks $\hat{\rho}_i = \text{rk}(\hat{\Gamma}_i)$ and largest eigenvalues $\lambda_1(\hat{\Gamma}_i)$ are asymptotically finite. Then

$$(1.1) \quad \frac{\|\hat{\Sigma}_1 - \hat{\Sigma}_2\|_F^2}{p} \leq \frac{\hat{\rho}_1 + \hat{\rho}_2}{p} \left[\lambda_1(\hat{\Gamma}_1) + \lambda_1(\hat{\Gamma}_2) \right]^2 \xrightarrow[n \rightarrow \infty]{} 0 \quad \text{a.s.}$$

This means we can interpret the two problems as asymptotically analogous in practice. This reasoning is short of being a formal result of equivalence, but will serve as a guiding principle.

We propose a solution to these parallel problems as follows. We first restrict ourselves to orthogonally invariant estimators of the spiked form; this large class can be thought as performing spiked corrections of the eigenvalues of the sample covariance matrix. For this class, there exists an unbiased risk estimator (URE) in the closely related invariant loss $L_H(\hat{\Sigma}, \Sigma) = \|\hat{\Sigma}\Sigma^{-1} - I\|_F^2/p$, which we will refer to as the Haff loss. We propose to find an optimal choice of noise estimator by minimizing this URE using calculus of variations. This approach is close in spirit to the work of [Stein \[1975, 1986\]](#), where he considers a loss based on a normal log-likelihood, although it is not specifically high-dimensional. It is also close to the Bayesian approach of [Haff \[1991\]](#). More generally, the idea to correct the eigenvalues of the sample covariance matrix is also found in previous work by [Ledoit and](#)

Wolf [2004], El Karoui [2008a], Ledoit and Wolf [2012] and Donoho et al. [2014].

The URE of the covariance estimator depends on first and second derivatives of the noise estimator, so directly minimizing the risk would yield an estimator that depends on the truth. It however happily turns out that the “dominant” part of this URE does not depend on the derivatives. It is therefore possible to obtain, in closed form, an estimator optimal for the dominant part of the URE.

We prove that our proposed estimator is well-behaved; for example, it is strongly consistent for σ^2 , even if the chosen estimators of γ_k and ρ are not. It is moreover essentially asymptotically normal of rate n , and we prove that this is the optimal minimax rate for the noise estimation problem. To illustrate concretely why this approach is interesting, we use it to construct a robust spiked covariance estimator. It seems to never perform worse than S in general, even in worst-case scenarios; while it performs remarkably well in spiked settings, and we show its eigenvalues are consistent.

We reiterate that in contrast with much work in high dimensional covariance estimation, we do not work with a sparsity assumption that many components of Σ or Σ^{-1} are zero. However, one can perfectly think of a spiked structure as a form of sparsity in itself, with ρ as sparsity parameter, which fits within the generally accepted principle that improved estimation in high dimensions is difficult unless some form of sparsity holds with the truth. The fact that we can construct an estimator that can exploit that structure when present, yet be robust to the assumption is encouraging.

The article is divided as follows. The regularity conditions, construction of the unbiased risk estimator and construction of the noise estimator is in Section 2. Investigations of properties of the noise estimator is done in Section 3. The example construction and simulations are in Section 4. After some comments in Section 5, we cover the proofs of the claims in Section 6. Notation. The following notation will be used throughout. We write $H_p(\mathbb{R})$ for the simplex $\{x \in \mathbb{R}^p \mid x_1 > \dots > x_p > 0\}$. The real p -dimensional orthogonal group is denoted $O_p(\mathbb{R})$. The Frobenius norm of a matrix A is the sum of its squared eigenvalues, denoted $\|A\|_F = \text{tr}(A^2)^{1/2}$, while the spectral norm is its largest singular value, denoted $\|A\|_2 = \sigma_{\max}(A)$. The notation $d_{\text{TV}}(\mu_1, \mu_2)$ stands for the total variation distance between two probability measures μ_1, μ_2 on an underlying measurable space (Ω, \mathcal{B}) , which equals $\sup_{A \in \mathcal{B}} |\mu_1(A) - \mu_2(A)|$. The p -dimensional Wishart distribution with n degrees of freedom and covariance matrix Σ is written $W_p(n, \Sigma)$.

2. CONSTRUCTION

We work in the following setting. Assume the data is an i.i.d. sample $X_1, \dots, X_n \sim N_p(0, \Sigma_p)$, with $n \geq p$ and $\Sigma_p > 0$. For such a sample, one can stack the data into a matrix $X = (X'_1, \dots, X'_n)$ and let $S = X'X/n = OLO'$, $L = \text{diag}(l_1, \dots, l_p)$ be the decreasing spectral decomposition of the sample

covariance matrix, with $l_1 > \dots > l_p > 0$ its ordered eigenvalues. The random matrix S , which is distributed as a scaled Wishart $n^{-1}W_p(n, \Sigma_p)$, serves as a naive estimator of Σ upon which we wish to improve. The normality and restriction to $n \geq p$ are necessary for the construction of the unbiased risk estimator that will follow; extensions will be discussed in Section 5.

As mentioned in the introduction, to adequately discuss high-dimensional behavior, we will also let this setting grow in complexity. We focus our attention on full-rank linear regimes, where a sequence of positive-definite covariance matrices of growing dimension $\Sigma_1, \Sigma_2, \Sigma_3, \dots$ is fixed; and $p = p_n$, as a function of the sample size, grows in the sense that $p_n/n \rightarrow c$ for some $c \in (0, 1)$. It will then be assumed that for every (n, p_n) , some i.i.d. sample $X_1, \dots, X_n \sim N_{p_n}(0, \Sigma_{p_n})$ will be available and a corresponding sample covariance matrix S constructed.

For such settings, the sequence $\{\Sigma_p\}$ is completely arbitrary beyond the requirement that each member be positive-definite. Of particular interest to us is the case where the covariance matrices form a *spiked sequence*, which we define as follows.

Definition 1. A sequence of covariance matrices $\{\Sigma_p\}$ is spiked if there exists a collection $\gamma_1 > \dots > \gamma_\rho > 0$ of size $\rho \geq 0$ and a $\sigma^2 > 0$ such that for any p , $\Sigma_p = \text{diag}(\gamma, 0) + \sigma^2 I_p$, where $\gamma = (\gamma_1, \dots, \gamma_\rho)$.

When discussing asymptotics, we will sometimes need that the spiked eigenvalues $\gamma_1, \dots, \gamma_\rho$ be sufficiently large with respect to the noise for efficient estimation to be possible. In practice, this will mean requiring that $\gamma_\rho/\sigma^2 > \sqrt{c}$, for c the asymptotic p_n/n ratio. The importance of this supercriticality condition for spiked eigenvalue estimation was first remarked by Baik et al. [2005] before being extended to the setting we are considering by Baik and Silverstein [2006], Paul [2007] and Nadler [2008]. They showed that for $1 \leq k \leq \rho$, the eigenvalues of the sample covariance matrix satisfy

$$(2.1) \quad l_k \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \begin{cases} \left(1 + c\sigma^2 \frac{\gamma_k + \sigma^2}{\gamma_k}\right) [\gamma_k + \sigma^2] & \text{if } \gamma_k > \sqrt{c}\sigma^2 \\ (1 + \sqrt{c})^2 \sigma^2 & \text{if } \gamma_k \leq \sqrt{c}\sigma^2 \end{cases},$$

with the $l_{\rho+1}, \dots, l_p$ asymptotically distributed like a scaled Marčenko-Pastur $\sigma^2 \text{MP}(c)$ distribution. Therefore, the asymptotic spectrum of S do not contain any information about those γ_k below the critical threshold $\sqrt{c}\sigma^2$. But since their estimation is mostly tangential to our goals, supercriticality will not always be necessary, and we will make it clear when it will be.

Let us now turn our attention to the task at hand. The parallel problems we wish to solve are

- (i) the estimation of σ^2 under the absolute error loss $L(\hat{\sigma}^2, \sigma^2) = |\hat{\sigma}^2 - \sigma^2|$;
- (ii) the estimation of Σ under the Frobenius loss $L_F(\hat{\Sigma}, \Sigma) = \|\hat{\Sigma} - \Sigma\|_F^2/p$ using spiked estimators.

Under spikedness, these two problems are parallel to each other in the sense of (1.1). The approach we take begins with aspect (ii) - we seek a good covariance estimator $\hat{\Sigma}$ in spiked form $\hat{\Gamma} + \hat{\sigma}^2 I$, with $\hat{\rho} = \text{rk}(\hat{\Gamma})$ small with respect to p , which we interpret as $\hat{\rho}$ a.s. tending to a finite constant. By appealing again to (1.1), it is clear that for the Frobenius loss, the spiked part $\hat{\Gamma}$ is asymptotically dominated by noise estimation. We therefore might as well *choose* $\hat{\Gamma}$ based on convenience: for example, we can pick one with consistent eigenvalue estimators, or some other property. A specific choice will be considered in Section 4. Alternatively, the recent results of Donoho et al. [2014] could provide an attractive choice based on consideration of single eigenvalue corrections, and we comment on this further in Section 5. Once a choice of $\hat{\Gamma}$ is made, we can then look for an optimal $\hat{\sigma}^2$, which would simultaneously solve aspects (i) and (ii) of the problem, while being asymptotically independent of our choice of $\hat{\Gamma}$.

Being quite free in selecting the spiked part, let us focus on mathematically convenient possibilities. A first restriction is to take $\hat{\Gamma}$ orthogonally invariant - that is, of the form $\hat{\Gamma} = O \text{diag}(\hat{\gamma}, 0) O'$ for some estimators $\hat{\gamma}_1 > \dots > \hat{\gamma}_{\hat{\rho}} > 0$ and $O \in O_p(\mathbb{R})$ the matrix of ordered eigenvectors of S . With this choice, our estimators $\hat{\Sigma}$ can be thought as performing spiked corrections on the sample covariance matrix S .

A second restriction will be necessary. At this point in our discussion the spiked, rank and noise estimators $\hat{\gamma}$, $\hat{\rho}$, $\hat{\sigma}^2$ have been essentially arbitrary. This is too general for the construction of the URE that will follow, so we must restrict ourselves to sufficiently regular estimators. The regularity conditions come in two flavors, weak and strong, and are statements of integrability; these conditions simply guarantee that expected values appearing in the construction of the URE are convergent. Combining the invariance and regularity restrictions, we define the following.

Definition 2. A spiked eigenvalue estimator $\hat{\Gamma}$ satisfies the weak regularity conditions if it is of the form $\hat{\Gamma} = O \text{diag}(\hat{\gamma}) O'$ for $S = OLO'$ and satisfies the following. Let $\hat{\rho} = \text{rk}(\hat{\Gamma})$. For each $1 \leq k \leq p$, $\hat{\gamma}_k$ are a.s. $C^2(H_p(\mathbb{R}); \mathbb{R})$ functions of l_1, \dots, l_p with boundary cases $\mathbb{1}[\hat{\rho} < k] \hat{\gamma}_k = 0$ and $\mathbb{1}[\hat{\rho} = p] \hat{\gamma}_k = \mathbb{1}[\hat{\rho} = p] l_k$ for which both expectations

$$\mathbb{E} \left[\left| \frac{\hat{\gamma}_k}{l_k} \right|^{9(1+\epsilon)} \right] \quad \text{and} \quad \mathbb{E} \left[\left| \frac{\partial \hat{\gamma}_k}{\partial l_k} \right|^{4.5} \right]$$

are finite for some $\epsilon > 0$. Similarly, a noise estimator $\hat{\sigma}^2$ satisfies the weak regularity conditions for a weak spiked eigenvalue estimator $\hat{\Gamma}$ if it is a $C^2(H_p(\mathbb{R}); \mathbb{R})$ function based on l_1, \dots, l_p such that for each $1 \leq k \leq p$,

$$\mathbb{E} \left[\left| \frac{\hat{\sigma}^2}{l_k} \right|^{9(1+\epsilon)} \right], \mathbb{E} \left[\left| \frac{\partial \hat{\sigma}^2}{\partial l_k} \right|^{4.5} \right] \quad \text{and} \quad \mathbb{E} \left[\left| \hat{\gamma}_k + \hat{\sigma}^2 \right| \left| \frac{\partial^2 \hat{\gamma}_k}{\partial l_k^2} + \frac{\partial^2 \hat{\sigma}^2}{\partial l_k^2} \right| \right]$$

are all finite for some $\epsilon > 0$.

The previous conditions assert integrability of quantities associated with the estimators for a given p , and are dimension dependent. In contrast, the following conditions assert similar integrability as p grows. To make the dependence explicit, we superscript the dimension.

Definition 3. A spiked eigenvalue estimator $\hat{\Gamma}^p$ satisfies the strong regularity conditions if it satisfies the weak regularity conditions for each $p > 0$, and moreover

$$\begin{aligned} & \mathbb{E} \left[\sup_{p>0} \max_{1 \leq k \leq p} \left| \frac{\hat{\gamma}_k^p}{l_k^p} \right|^{9(1+\epsilon)} \right], \mathbb{E} \left[\sup_{p>0} \max_{1 \leq k \leq p} \left| \frac{\partial \hat{\gamma}_k^p}{\partial l_k^p} \right|^{4.5} \right], \mathbb{E} \left[\sup_{p>0} \max_{1 \leq k \leq p} \left| \hat{\gamma}_k^p \frac{\partial^2 \hat{\gamma}_k^p}{\partial l_k^{p2}} \right| \right], \\ & \mathbb{E} \left[\sup_{p>0} \max_{1 \leq k \neq b \leq \hat{\rho}} \left| \frac{\hat{\gamma}_k^p - \hat{\gamma}_b^p}{l_k^p - l_b^p} \right|^2 \right], \mathbb{E} \left[\sup_{p>0} \max_{\substack{1 \leq k \neq b \\ \neq e \leq \hat{\rho}}} \left| \frac{l_k^p}{l_k^p - l_b^p} \right|^2 \left| \frac{\hat{\gamma}_k^p - \hat{\gamma}_e^p}{l_k^p - l_e^p} - \frac{\hat{\gamma}_b^p - \hat{\gamma}_e^p}{l_b^p - l_e^p} \right|^2 \right] \\ & \text{and } \mathbb{E} \left[\sup_{p>0} \max_{1 \leq k \neq b \leq \hat{\rho} < e \leq p} \left| \frac{l_k^p}{l_k^p - l_b^p} \right|^2 \left| \frac{\hat{\gamma}_k^p}{l_k^p - l_e^p} - \frac{\hat{\gamma}_b^p}{l_b^p - l_e^p} \right|^2 \right] \end{aligned}$$

are all finite for some $\epsilon > 0$. Similarly, a noise estimator $\hat{\sigma}^2$ satisfies the strong regularity conditions for a strong spiked eigenvalue estimator $\hat{\Gamma}$ if it satisfies the weak, and the following holds:

$$\begin{aligned} & \mathbb{E} \left[\sup_{p>0} \max_{0 \leq k \leq p} \left| \frac{\hat{\sigma}^{2p}}{l_k^p} \right|^{9(1+\epsilon)} \right], \mathbb{E} \left[\sup_{p>0} \max_{0 \leq k \leq p} \left| \frac{\partial \hat{\sigma}^{2p}}{\partial l_k^p} \right|^{4.5} \right] \\ & \text{and } \mathbb{E} \left[\sup_{p>0} \max_{1 \leq k \leq p} \left| \hat{\gamma}_k^p + \hat{\sigma}^{2p} \right| \left| \frac{\partial^2 \hat{\gamma}_k^p}{\partial l_k^{p2}} + \frac{\partial^2 \hat{\sigma}^{2p}}{\partial l_k^{p2}} \right| \right] \end{aligned}$$

are all finite for some $\epsilon > 0$.

Careful inspection of the proofs reveal that regularity conditions in this spirit are inevitable; however, we emphasize that by no means we believe those precise conditions to be necessary, merely sufficient. In any case, with these conditions in hand we can formally define the classes of estimators in which we look for an optimal $\hat{\sigma}^2$.

Definition 4. For $\hat{\Gamma}$ a weak (strong) spiked eigenvalue estimator, the associated weak (strong) class of spiked corrections to the sample covariance matrix is

$$V_p(\hat{\Gamma}) = \left\{ \hat{\Gamma} + \hat{\sigma}^2 I_p \mid \hat{\sigma}^2 \text{ is } \hat{\Gamma}\text{-weak} \right\} \text{ and } \bar{V}_p(\hat{\Gamma}) = \left\{ \hat{\Gamma} + \hat{\sigma}^2 I_p \mid \hat{\sigma}^2 \text{ is } \hat{\Gamma}\text{-strong} \right\}.$$

We would like to find an optimal estimator over these two classes. Recall we are evaluating performance in Frobenius loss $L_F(\hat{\Sigma}, \Sigma) = \|\hat{\Sigma} - \Sigma\|_F^2/p$. Although natural and common within the literature, we find it more convenient to move to the closely related “invariant” loss

$$L_H(\hat{\Sigma}, \Sigma) = \frac{\|\hat{\Sigma}\Sigma^{-1} - I\|_F^2}{p}.$$

This loss was, up to the high-dimensional p^{-1} normalization, mentioned by [James and Stein \[1961\]](#) early but first thoroughly investigated by [Haff \[1977\]](#), and we will refer to it as Haff's loss. A modification of the argument behind (1.1) shows that estimation of a spiked covariance matrix under this loss can also be thought as a noise estimation problem, just like for the Frobenius case. In this sense the problem stays similar.

A great advantage of the Haff loss is that it is one of the few for which an unbiased estimator of the risk is known, at least in the orthogonally invariant case. There is a rich body of literature behind that construction [[Haff, 1977, 1979, 1980](#)], in different shapes and under different conditions. A remarkable feature is that if we collect and split the terms of the URE between the terms of leading and smaller order, the dominant part does not depend on the derivatives of the eigenvalue estimators. More precisely, we have this construction.

Theorem 1. *Let $n \geq p + 1$. Then for any weak spiked estimator $\hat{\Sigma} \in V_p(\hat{\Gamma})$ whose spiked rank $\hat{\rho}$ is independent of S , we find its Haff risk to satisfy $E[L_H(\hat{\Sigma}, \Sigma)] = E[F + G]$ with $E[|F + G|] < \infty$, where*

$$F = F(l, \hat{\rho}, \hat{\gamma}, \hat{\sigma}^2) \quad \text{and} \quad G = G\left(l, \hat{\rho}, \hat{\gamma}, \hat{\sigma}^2, \frac{\partial \hat{\gamma}}{\partial l}, \frac{\partial \hat{\sigma}^2}{\partial l}, \frac{\partial^2 \hat{\gamma}}{\partial l^2}, \frac{\partial^2 \hat{\sigma}^2}{\partial l^2}\right)$$

are functionals that do not depend on Σ . In addition, if the estimator is strong in the sense that $\hat{\Sigma} \in \bar{V}_p(\hat{\Gamma})$, then asymptotically F is the dominant term and G the dominated term, respectively, in the sense that

$$\lim_{n \rightarrow \infty} E[|F|] < \infty \quad \text{and} \quad \lim_{n \rightarrow \infty} p E[|G|] < \infty.$$

Explicit expressions for F and G are given in (6.10)–(6.11).

In practice, we found the decomposition most useful for a deterministic spiked rank $\hat{\rho} = r$, in which case we might consider estimates of the form $\hat{\Gamma}_r + \sigma_r^2 I$; this is the approach we take in Section 4 when constructing a specific estimator. But it is reasonable to think of a context in which some estimate of the true rank ρ based on prior independent data is available, in which case the construction applies equally.

Now fix a weak spiked eigenvalue estimator $\hat{\Gamma}$, and consider the task of finding a $\hat{\sigma}^2$ that minimizes the Haff risk: minimizing the construction from Theorem 1 makes the task plausible. Since G depends on the derivatives of $\hat{\sigma}^2$, formally proceeding with calculus of variations would yield a minimizer that depends on the unknown density of the eigenvalues which is, of course, unavailable. However since the dominant part only depends on $\hat{\sigma}^2$ itself, one can obtain a minimizer of $E[F]$ whose expression is independent of Σ .

Proposition 1. *Let $n \geq p$. If $\hat{\Sigma} \mapsto E[F]$ has a minimum over $V_p(\hat{\Gamma})$, it is given by $\tilde{\Sigma} = \hat{\Gamma} + \tilde{\sigma}^2 I$, where $\tilde{\sigma}^2 = A/B$ where*

$$\begin{aligned}
A &= \frac{n-p-1}{np} \sum_{c=1}^p \frac{1}{l_c} - \frac{(n-p-1)(n-p-2)}{n^2 p} \sum_{k=1}^{\hat{\rho}} \frac{\hat{\gamma}_k}{l_k^2} \\
&\quad + \frac{n-p-1}{n^2 p} \sum_{k=1}^{\hat{\rho}} \sum_{c=1}^p \frac{\hat{\gamma}_k}{l_k} \frac{1}{l_c} - 2 \frac{n-p-1}{n^2 p} \sum_{k=1}^{\hat{\rho}} \sum_{c=\hat{\rho}+1}^p \frac{1}{l_c} \frac{\hat{\gamma}_k}{l_k - l_c} \\
&\quad + \frac{3}{n^2 p} \sum_{k \neq b}^{\hat{\rho}} \sum_{c=\hat{\rho}+1}^p \frac{1}{l_k - l_c} \frac{\hat{\gamma}_b}{l_b - l_c} - \frac{3}{n^2 p} \sum_{k=1}^{\hat{\rho}} \sum_{c \neq d=\hat{\rho}+1}^p \frac{1}{l_k - l_c} \frac{\hat{\gamma}_k}{l_k - l_d} \\
&\quad - \frac{3}{n^2 p} \sum_{k \neq b}^{\hat{\rho}} \sum_{c=\hat{\rho}+1}^p \frac{\hat{\gamma}_k - \hat{\gamma}_b}{l_k - l_b} \frac{1}{l_k - l_c}, \\
B &= \frac{(n-p-1)(n-p-2)}{n^2 p} \sum_{c=1}^p \frac{1}{l_c^2} - \frac{n-p-1}{n^2 p} \sum_{c=1}^p \frac{1}{l_c} \sum_{c=1}^p \frac{1}{l_c}.
\end{aligned}$$

In addition, if $n \geq 2p + 2$, a minimum must exist (and therefore at $\tilde{\Sigma}$).

The matter of whether a minimizer should exist at all in a given context is delicate. A proof of existence for some large class of covariance matrices would be quite interesting. In the spiked case, the remarks following Lemma 1 hint at a plausible approach.

3. PROPERTIES

The previous chapter was concerned with the construction of a good estimator $\tilde{\sigma}^2$ that satisfies some optimality property, namely minimizing the dominant part of the Haff URE over $V_p(\hat{\Gamma})$. Let us now turn our attention to its performance in estimating σ^2 under spikedness. We will make repeated use of the following lemma, which extends the results of Nadler [2008].

Lemma 1. *Suppose the underlying sequence of covariance matrices $\{\Sigma_p\}$ is spiked and $p_n/n \rightarrow c \in (0, 1)$.*

(i) *If $\gamma_\rho/\sigma^2 > \sqrt{c}$, then for any $1 \leq k \leq \rho$,*

$$\frac{1}{p-\rho} \sum_{c=\rho+1}^p \frac{l_c}{l_k - l_c} \xrightarrow[n \rightarrow \infty]{a.s.} \frac{\sigma^2}{\gamma_k};$$

(ii) *For any $m > 1$,*

$$\frac{1}{p-\rho} \sum_{c=\rho+1}^p \frac{1}{l_c^m} \xrightarrow[n \rightarrow \infty]{a.s.} \frac{1}{(1-c)^{2m-1}} \frac{1}{\sigma^{2m}}.$$

The supercriticality assumption $\gamma_\rho/\sigma^2 > \sqrt{c}$ in (i) is necessary for the expression to converge. Two remarks are in order. First, as a consequence

of this result, it is easy to show that

$$\begin{aligned} \frac{1}{p-\rho} \sum_{c=\rho+1}^p \frac{1}{l_k - l_c} &\xrightarrow[n \rightarrow \infty]{\text{a.s.}} \frac{1}{\gamma_k + c\sigma^2}, \\ \frac{1}{p-\rho} \sum_{c=\rho+1}^p \frac{1}{l_c(l_k - l_c)} &\xrightarrow[n \rightarrow \infty]{\text{a.s.}} \frac{1}{1-c} \frac{\gamma_k}{\sigma^2(\gamma_k + c\sigma^2)^2}, \end{aligned}$$

a result we will use later in Section 4. Second, in connection with the proof of Proposition 1, we see that when the underlying sequence of covariance matrices $\{\Sigma_p\}$ is spiked

$$\frac{1}{n^2} \left[(n-p-2) \sum_{c=1}^p \frac{1}{l_c^2} - \left(\sum_{c=1}^p \frac{1}{l_c} \right)^2 \right] \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \frac{c}{(1-c)\sigma^4} > 0,$$

by Lemma 1. From (6.13), one would therefore expect the estimator also to be a minimizer under spikedness. Although we haven't been successful in formalizing this intuition, this could be a plausible approach towards proving existence of minimizers for spiked covariance matrices.

Let us now turn our attention to the behavior of $\tilde{\sigma}^2$. The following theorem summarizes important aspects of its asymptotic behavior.

Theorem 2. *Suppose the underlying sequence of covariance matrices $\{\Sigma_p\}$ is spiked and $p_n/n \rightarrow c \in (0, 1)$ with $\gamma_\rho/\sigma^2 > \sqrt{c}$. For a given weak $\hat{\Gamma}$, let $\tilde{\sigma}^2$ be the associated minimizer of Proposition 1. Then*

- (i) *If $\hat{\rho}$ a.s. converges to a finite constant, then $\tilde{\sigma}^2 \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \sigma^2$.*
- (ii) *If $\hat{\rho}$ is strongly consistent and for all $1 \leq k \leq \rho$, $\hat{\gamma}_k$ a.s. converges to some finite constant, then we have bounds $X_n^- \leq n(\tilde{\sigma}^2 - \sigma^2) \leq X_n^+$ with*

$$X_n^- \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N\left(\mu^-, \frac{2c(1+c)^2}{(1-c)^4} \sigma^4\right), \quad X_n^+ \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N\left(\mu^+, \frac{2c(1+c)^2}{(1-c)^4} \sigma^4\right),$$

where μ^- and μ^+ have explicit expressions given in (6.25)–(6.26).

An immediate consequence of this result is that $\tilde{\sigma}^2$ estimates σ^2 with rate n in our absolute error loss $L(\hat{\sigma}^2, \sigma^2) = |\hat{\sigma}^2 - \sigma^2|$. We should mention that, given good estimators of ρ and γ_k , one could perhaps build approximate high-dimensional confidence intervals for σ^2 with part (ii) of Theorem 2. We will not investigate this further, but rather turn our attention to minimax rates for the noise estimation problem. For any spiked sequence $\{\Sigma_p\}$, define a δ -ball of order r as

$$B_r(\Sigma, \delta) = \left\{ \{\Sigma'_p\} \text{ spiked} \mid \left| \lambda_i(\Sigma_p) - \lambda_i(\Sigma'_p) \right| < \delta \frac{\lambda_i(\Sigma_p)}{n^r} \quad \forall p > 0 \right\}.$$

We start with a lemma. Recall that d_{TV} stands for the total variation distance between two probability measures.

Lemma 2. *Let $\{\Sigma_p\}$ be a spiked sequence of covariance matrices and $M > 0$. Then, as $p_n/n \rightarrow c \in (0, 1)$,*

$$\lim_{n \rightarrow \infty} \sup_{\Sigma' \in B_r(\Sigma, 2M)} d_{TV} \left(N(0, \Sigma_p)^n, N(0, \Sigma'_p)^n \right) \leq \sqrt{1 - \exp \left(-\frac{cM^2}{2} \right)}$$

when $r = 1$, while this limit is zero when $r > 1$.

Using this result, we now proceed to show a lower bound on the local minimax rate of convergence for estimating σ^2 , using the classic two-point test argument of [Le Cam \[1973\]](#).

Theorem 3. *Say the underlying sequence of covariance matrices $\{\Sigma_p\}$ is spiked and $p_n/n \rightarrow c \in (0, 1)$. Let $\epsilon > 0$ and define*

$$M_\epsilon = \sqrt{-\frac{2}{c} \log \left(1 - (1 - 4\epsilon)^2 \right)}.$$

Then no estimator can estimate σ^2 with speed $\sigma^2 M_\epsilon/n$ over the shrinking neighborhoods $B(\Sigma_p, 2M_\epsilon)$, in the sense that

$$\liminf_{n \rightarrow \infty} \inf_{\hat{\sigma}^2} \sup_{\Sigma' \in B_1(\Sigma, 2M_\epsilon)} P_{\Sigma'_p} \left[|\hat{\sigma}^2 - \sigma^{2'}| > \sigma^2 \frac{M_\epsilon}{n} \right] \geq \epsilon.$$

Thus, the minimax rate of estimation of σ^2 over n -shrinking neighborhoods cannot be faster than $O_P(1/n)$ (so in particular over, say, fixed neighborhoods.) Using [Theorem 2](#), we can show our noise estimator $\hat{\sigma}^2$ essentially achieves this rate, in the sense that it is $o_P(1/n^r)$ over n^r -shrinking neighborhoods for any $r > 1$.

Proposition 2. *Let the underlying sequence of covariance matrices $\{\Sigma_p\}$ be spiked and $p_n/n \rightarrow c \in (0, 1)$ with $\gamma_\rho/\sigma^2 > \sqrt{c}$. For a given weak $\hat{\Gamma}$, let $\tilde{\sigma}^2$ be the associated extremizer of [proposition 1](#). If $\hat{\rho}$ is strongly consistent and for all $1 \leq k \leq \rho$, $\hat{\gamma}_k$ a.s. converges to some finite constant, then for any $r > 1$ and $M > 0$, $\tilde{\sigma}^2$ estimates σ^2 with rate at least $\sigma^2 M/n^r$ over the shrinking neighborhoods $B_r(\Sigma, 2M)$, in the sense that*

$$\lim_{n \rightarrow \infty} \sup_{\Sigma' \in B_r(\Sigma, 2M)} P_{\Sigma'_p} \left[|\tilde{\sigma}^2 - \sigma^{2'}| > \sigma^2 \frac{M}{n^r} \right] = 0.$$

Thus we can conclude that, despite choosing our noise estimator to minimize a covariance problem, good behavior has been transferred to the noise estimation problem, which is not surprising in light of [\(1.1\)](#). In particular, we see that by [Theorem 2 \(i\)](#), strong consistency of the noise estimator follows even when the eigenvalues of $\hat{\Gamma}$ itself are not consistent – a robustness which is certainly welcome.

4. APPLICATION

Having built and analyzed our noise estimator, we now proceed to illustrate our construction by building a specific covariance estimator. We hope this concrete example will help clarify the approach taken and its behavior in the covariance problem.

4.1. Example. We build a spiked covariance estimator as follows. The first step is to specify an asymptotically negligible spiked component $\hat{\Gamma}$. For r be some fixed rank strictly smaller than p , we take $\tilde{\Gamma}_r = O \text{diag}(\tilde{\gamma}) O'$ with

$$\tilde{\gamma}_k = \sum_{c=r+1}^p l_c \left(\sum_{c=r+1}^p \frac{l_c}{l_k - l_c} \right)^{-1}$$

for $1 \leq k \leq r$, and 0 otherwise. These estimators are strongly consistent when $r = \rho$, as we will soon show using Lemma 1; this is the main motivation for our choice. Note that this choice does not quite fit within the framework considered by Donoho et al. [2014], since it is not a function of l_k only. With this choice of spiked part, let $\tilde{\sigma}_r^2$ be the minimizer of Proposition 1 associated with our spiked component. We then have a family $r \rightarrow \tilde{\Sigma}_r = \tilde{\Gamma}_r + \tilde{\sigma}_r^2 I$ for all $0 \leq r < p$, which we naturally extend to the $r = p$ case through $\tilde{\Sigma}_p = \tilde{\Gamma}_p = S$.

Next, we select the rank r based on the data. Motivated again by the results of Lemma 1, we define the rank estimator

$$\tilde{\rho} = \arg \min_{0 \leq r \leq p} \left\{ \frac{\mathbb{1}[r < p]}{l_{r+1}} \frac{(1 + \sqrt{p/n})^2}{p - r} \sum_{c=r+1}^p l_c \geq 1, \quad |F_r + G_r| \leq \frac{p+1}{n} \right\},$$

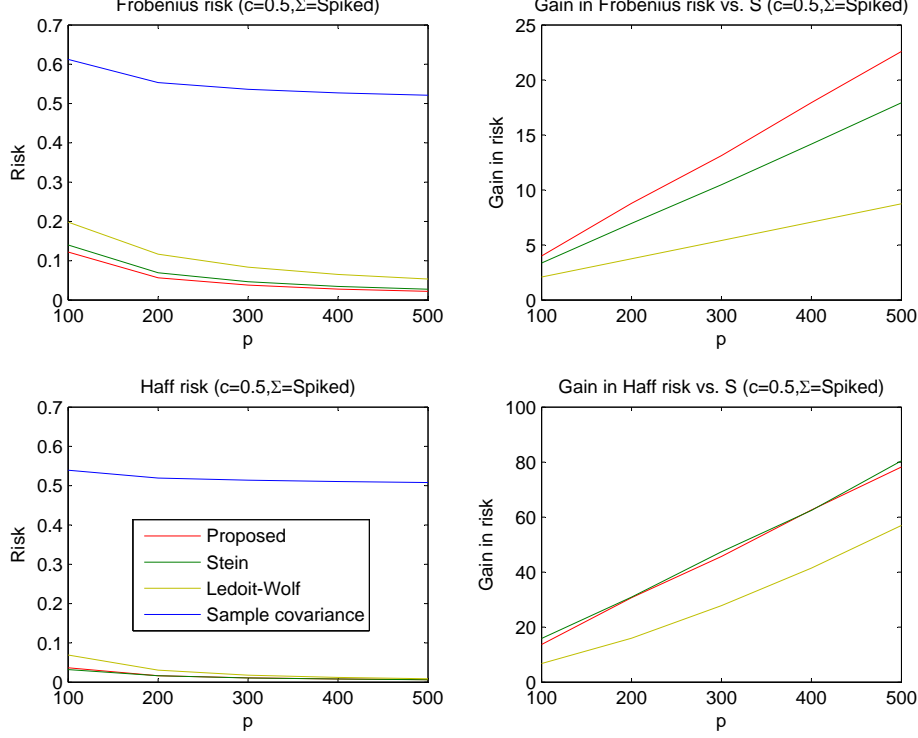
with F_r, G_r the F, G of Theorem 1 applied to $\tilde{\Gamma}_r$ and $\tilde{\sigma}_r^2$. This choice aims to select the smallest rank that lies both above the critical threshold and yields improvement in Haff risk over S . Since $r = p$ satisfies both criteria, the set is never empty and in the worst case we simply do not correct the eigenvalues of S . This will happen when there is strong departure from spikedness, which means that the construction is in some sense robust to this situation: it exploits it when present and reverts to S when not.

Finally, we simply set $\tilde{\Sigma} = \tilde{\Sigma}_{\tilde{\rho}}$ as our estimator. In practice, the computation is straightforward, since everything is in closed-form, with polynomial complexity. An implementation is available as an R package at <http://stat.cornell.edu/~chetelat>. At the same time, as previously hinted, the estimator has strongly consistent eigenvalues under spikedness. The proof is a simple application of results from Section 3.

Proposition 3. *If the underlying sequence of covariance matrices $\{\Sigma_p\}$ is spiked and $p_n/n \rightarrow c \in (0, 1)$ with $\gamma_\rho/\sigma^2 > \sqrt{c}$, then $\tilde{\rho}$, $\tilde{\sigma}^2$ and $\tilde{\gamma}_k$ for $1 \leq k \leq \rho$ are all strongly consistent.*

This stands, of course, in contrast with the eigenvalues of the sample covariance matrix S , which converge to the wrong values (2.1) in spiked settings.

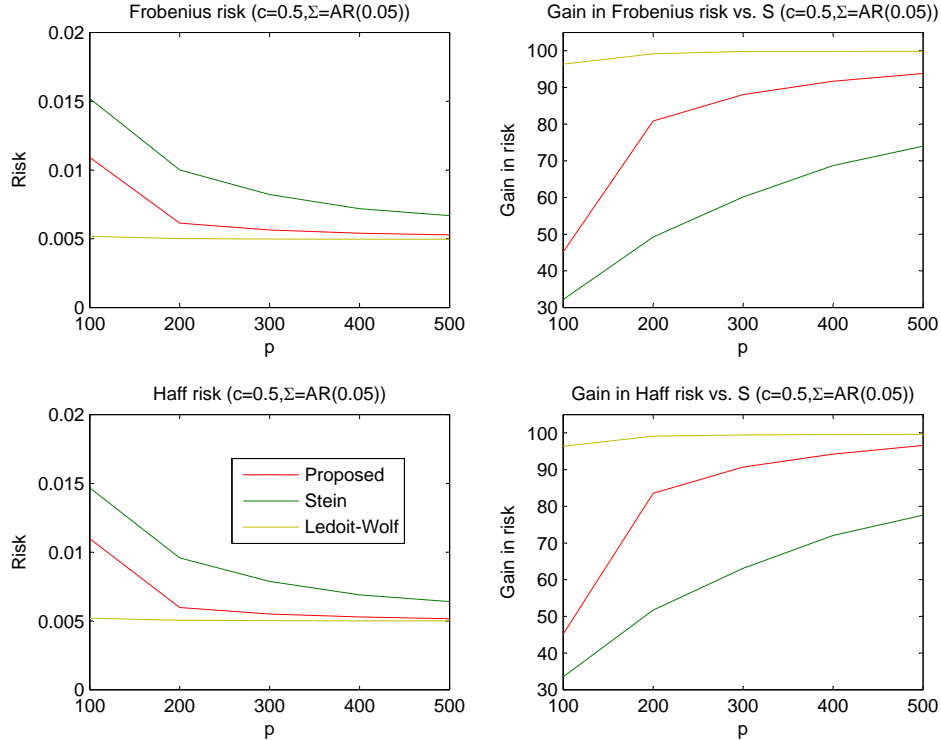
FIGURE 1. Spiked covariance setting.



4.2. Numerical Comparisons. We display the performance of the constructed estimator through simulations. The setting is as follows. We fix the dimension sample ratio c at 0.5 and vary n, p . For each n, p , we simulate data from a normal $N(0, \Sigma)$ and approximate its Haff and Frobenius risk using a law of large numbers approximation with 100 iterations. Four true covariance matrices Σ are considered. The first is a spiked setting $\Sigma = \text{diag}(5, 4, 3, 2, 1, \dots, 1)$, while the other three correspond to autoregressive settings $\Sigma_{ij} = \kappa^{|i-j|}$ for $\kappa = 0.05, 0.5$ and 0.95 . The case $\kappa = 0.95$ is particularly difficult for the constructed estimator as it is very far from spikedness.

The risks are computed for S , our estimator $\tilde{\Sigma}$ and two benchmark competitors. The first is Stein's isotonized covariance estimator, with well regarded overall performance. We follow the implementation of [Lin and Perlman \[1985\]](#). The second is the popular linear shrinkage covariance estimator of [Ledoit and Wolf \[2004\]](#), specifically designed for high-dimensional settings. We plot the risks and the gain in risk with respect to S , defined as $\text{Risk}(S)/\text{Risk}(\tilde{\Sigma}) - 1$. The computations were performed using the R package, and the results are given as Figures 1-4. Blue corresponds to S , red to $\tilde{\Sigma}$, green to Stein's isotonized estimator and yellow to the Ledoit-Wolf estimator.

FIGURE 2. AR(0.05) setting. The sample covariance matrix is omitted.

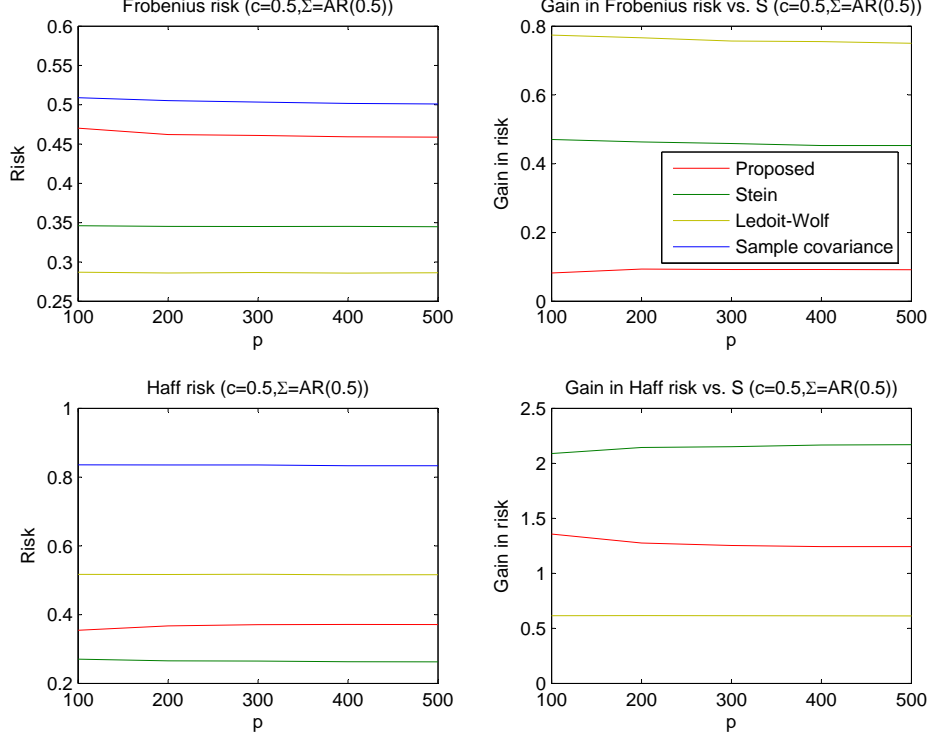


We can see from the results of the simulations that the expected good performance in Haff loss of our estimator seems to translate well into the more standard Frobenius loss. The estimator performs particularly well in supercritical spiked settings, with a 23-times improvement over S in Frobenius risk for $p = 500$, $n = 1000$ in our setting. In addition, the estimator is quite robust to deviations from spikedness, as even in worst-case scenarios such as an AR(0.95) setting, we do not do worse than S in Haff or Frobenius risk. There is therefore little to lose by using it rather than the sample covariance matrix, and as far as such a thing can exist, it could be advocated as some kind of generic high-dimensional covariance estimator.

5. COMMENTS

In this work we considered two parallel high-dimensional problems, the estimation of noise in principal components analysis under absolute error loss and the estimation of a spiked covariance matrix under Frobenius loss. We proposed a variational solution, by restricting ourselves to regular estimators and minimizing an unbiased covariance risk estimator in the invariant analogue of the loss. The resulting noise estimator was shown to be strongly consistent and almost asymptotically normal and minimax for the

FIGURE 3. AR(0.50) setting.

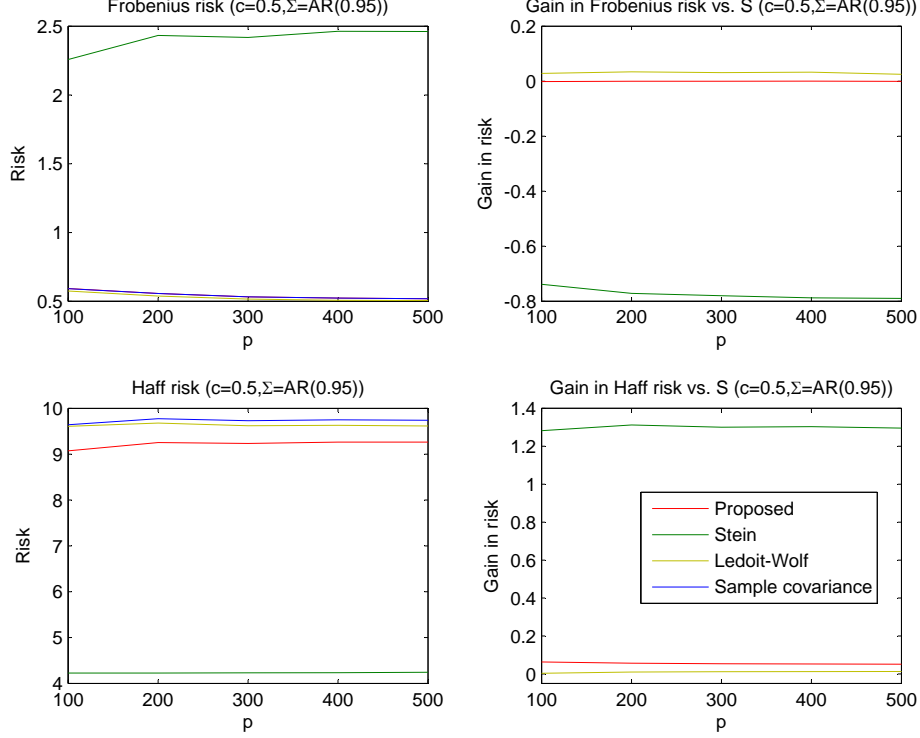


noise problem, and we used the construction to build a robust spiked covariance estimator with good simulation performance. Beyond this, however, there are several aspects of our solution that warrant further discussion.

First, we assumed throughout this work that the underlying data was normal, and the construction and proofs depend quite heavily on it. This could be a point of discord between practice and the theory outlined here. However, we feel that, unlike many statistical problems where normality is convenient but unrealistic, it is quite natural here. Indeed, the construction and its properties only depends on the data through the eigenstructure of S , unlike other estimators such as the one of [Ledoit and Wolf \[2004\]](#). The sample covariance matrix being an empirical average, one can expect it to behave asymptotically like a Wishart, and in that sense the assumption does not appear particularly restrictive.

Another assumption running through the work is that although we are in high-dimensions, we keep $p \leq n$. The extension to a $p > n$ setting is attractive, as in addition to the properties described above, a corresponding robust spiked covariance estimator would be automatically invertible, in contrast with S . The single obstacle appears to be the absence of an appropriate unbiased risk estimator for the Haff loss when $p > n$. This is therefore more an obstruction by knowledge than mathematics, as if such a

FIGURE 4. AR(0.95) setting.



construction would be found, the method outlined in this work could easily be applied.

As we considered minimization of a covariance loss, it might be surprising that we did not present any results on the behavior of the estimator in the covariance problem. We strongly believe that the Haff risk must tend to zero under spikedness since, as some algebra shows, the unbiased risk estimator of our estimate tends a.s. to zero. This is quite interesting since the Haff risk of S equals, in contrast, $(p+1)/n \rightarrow c > 0$. Unfortunately, we haven't been able to prove this statement. Although the literature on the probabilistic behavior of Wishart eigenvalues is extensive, it is more scant on their L^1 behavior, and this limits what can be proven as of now.

In Section 2, we considered an invariant analogue of the Frobenius loss $\|\hat{\Sigma} - \Sigma\|_F^2/p$, the Haff loss $\|\hat{\Sigma}\Sigma^{-1} - I\|_F^2/p$ which allowed for the existence of an unbiased risk estimator. Since our estimator is particularly adapted to this invariant covariance loss, it might also be of interest to study an invariant noise loss such as $|\hat{\sigma}^2/\sigma^2 - 1|$.

We did not tackle the problem of selecting the spiked eigenvalue estimators $\hat{\gamma}_k$ in an optimal way, beyond the suggestion in Section 4. The recent work of Donoho et al. [2014] could offer a solution. The authors consider the spiked

covariance estimation problem where the noise is known and fixed at $\sigma^2 = 1$, and look for optimal shrinking of the spiked eigenvalues l_k , $1 \leq k \leq \rho$. In the Frobenius and Haff losses, their optimal estimators coincide and equal

$$\hat{\gamma}_k = \left[l_k - 1 + \frac{cl_k}{l_k - 1} \right] \frac{1 - c/(l_k - 1)^2}{1 + c/(l_k - 1)}$$

for $l_k > (1 + \sqrt{c})^2$. An appealing feature of this estimate is that it accounts for the deterministic angles between the top sample and population eigenvectors. It would be interesting to study the behavior of the noise estimator $\tilde{\sigma}^2$ from Theorem 1 applied to these spiked estimators, with perhaps adjustments for not knowing σ^2 .

Finally, we should remark that our construction automatically provides well-conditioned covariance estimators, which is quite important for applications. Therefore, when the parameter of interest is the precision rather than the covariance matrix, using $\tilde{\Sigma}^{-1}$ as estimator appears reasonable, although we currently do not have any formal results on its behavior for this problem.

6. TECHNICAL RESULTS AND PROOFS

6.1. Proofs for Section 2. The following Stein-Haff identity is used to compute an unbiased estimator of risk for orthogonally invariant estimators in proposition 1. The general identity dates back to Haff [1979] and its specialization to orthogonally invariant estimators for $n \geq p$ first implicitly used by Sheena [1995]. Unfortunately, the approach taken by the author requires regularity conditions that are difficult to verify in practice (conditions 1–3 in his Section 1 and 2). The following lemma follows the approach of Konno [2009] and Kubokawa and Srivastava [2008] to obtain the same $n \geq p$ result, but under weaker, simpler conditions.

Lemma 3. *Let $W \sim W_p(n, \Sigma)$ with $n \geq p$, and let $W/n = OLO'$ be the spectral decomposition of the associated sample covariance matrix. Let $\psi_1(L), \dots, \psi_p(L)$ be differentiable functions of the eigenvalues of W/n satisfying:*

$$(6.1) \quad \mathbb{E} \left[\left| \sum_{k=1}^p \frac{n-p-1}{n} \frac{\psi_k}{l_k} + \frac{2}{n} \sum_{k=1}^p \frac{\partial \psi_k}{\partial l_k} + \frac{1}{n} \sum_{k \neq b}^p \frac{\psi_k - \psi_b}{l_k - l_b} \right| \right] < \infty.$$

Define $\Psi = \text{diag}(\psi_1, \dots, \psi_p)$. Then

$$\mathbb{E}[\text{tr}(\Sigma^{-1} O \Psi O')] = \mathbb{E} \left[\sum_{k=1}^p \frac{n-p-1}{n} \frac{\psi_k}{l_k} + \frac{2}{n} \sum_{k=1}^p \frac{\partial \psi_k}{\partial l_k} + \frac{1}{n} \sum_{k \neq b}^p \frac{\psi_k - \psi_b}{l_k - l_b} \right].$$

Proof. We use Lemma 3 in Chételat and Wells [2012]. Decompose $W = X'X$ for some $X \sim N_{n \times p}(0, I_n \otimes \Sigma)$. In the spirit of Lemma 4.1 in Konno

[2009], we find

$$\begin{aligned} (dX')X + X'dX &= d(X'X) = n(dO)LO' + nOdLO' + nOLdO' \\ \Rightarrow O' [(dX')X + X'dX] O &= nO'dOL + nL(dO')O + ndL. \end{aligned}$$

Since $O'dO + (dO')O = 0$, we get

$$O' [(dX')X + X'dX] O = nO'dOL - nLO'dO + ndL,$$

so that for $k \neq l$,

$$\begin{aligned} O'dO_{kl} &= \frac{1}{n} \frac{1}{l_l - l_k} O' [(dX')X + X'dX] O_{kl}, \\ dL_{kk} &= \frac{1}{n} O' [(dX')X + X'dX] O_{kk} \end{aligned}$$

and $O'dO_{kk} = 0$. Then

$$\begin{aligned} \frac{\partial l_k}{\partial X_{ij}} &= \frac{1}{n} \sum_{\alpha, \beta, \gamma} O'_{k\alpha} \frac{X'_{\alpha\beta}}{dX_{ij}} X_{\beta\gamma} O_{\gamma k} + \frac{1}{n} \sum_{\alpha, \beta, \gamma} O'_{k\alpha} X'_{\alpha\beta} \frac{X_{\beta\gamma}}{dX_{ij}} O_{\gamma k} \\ (6.2) \quad &= \frac{2}{n} \sum_{\gamma} O'_{kj} X_{i\gamma} O_{\gamma k} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial O_{kl}}{\partial X_{ij}} &= \frac{1}{n} \sum_{\alpha \neq l, \beta, \gamma, \epsilon} O_{k\alpha} \frac{1}{l_l - l_\alpha} O'_{\alpha\beta} \left[\frac{\partial X'_{\beta\gamma}}{\partial X_{ij}} X_{\gamma\epsilon} + X'_{\beta\gamma} \frac{\partial X_{\gamma\epsilon}}{\partial X_{ij}} \right] O_{\epsilon l} \\ (6.3) \quad &= \frac{1}{n} \sum_{\alpha \neq l, \beta} O_{k\alpha} \frac{O'_{\alpha j} O_{\beta l} + O'_{\alpha\beta} O_{jl}}{l_l - l_\alpha} X_{i\beta}. \end{aligned}$$

Now define $\tilde{X} = X\Sigma^{-1/2}$ and $H = \frac{1}{n}\Sigma^{1/2}OL^{-1}\Psi O'\Sigma^{-1/2}$ – we need to compute $\text{div}_{\text{vec}(\tilde{X})} \text{vec}(\tilde{X}H)$. We find

$$\begin{aligned} \text{div}_{\text{vec}(\tilde{X})} \text{vec}(\tilde{X}H) &= \sum_{\alpha, i, j} \frac{\partial}{\partial \tilde{X}_{\alpha i}} \left\{ \tilde{X}_{\alpha j} H_{ji} \right\} = n \sum_i H_{ii} + \sum_{\alpha, j} \tilde{X}_{\alpha j} \frac{\partial H_{ji}}{\partial \tilde{X}_{\alpha i}} \\ &= \sum_{\gamma} \frac{\psi_{\gamma}}{l_{\gamma}} + \frac{1}{n} \sum_{\alpha, \beta, i, j, k, l} \tilde{X}_{\alpha j} \Sigma_{\beta i}^{1/2} \Sigma_{jk}^{1/2} \frac{\partial}{\partial X_{\alpha\beta}} \{OL^{-1}\Psi O'_{kl}\} \Sigma_{li}^{-1/2} \\ (6.4) \quad &= \sum_{\gamma} \frac{\psi_{\gamma}}{l_{\gamma}} + \frac{1}{n} \sum_{\alpha, k, l} X_{\alpha k} \frac{\partial}{\partial X_{\alpha l}} \{OL^{-1}\Psi O'_{kl}\} \\ &= \sum_{\gamma} \frac{\psi_{\gamma}}{l_{\gamma}} + \frac{1}{n} \sum_{\alpha, k, l, \beta} X_{\alpha k} \frac{\partial O_{k\beta}}{\partial X_{\alpha l}} [L^{-1}\Psi]_{\beta\beta} O'_{\beta l} \\ &\quad + \frac{1}{n} \sum_{\alpha, k, l, \beta} X_{\alpha k} O_{k\beta} \frac{\partial [L^{-1}\Psi]_{\beta\beta}}{\partial X_{\alpha l}} O'_{\beta l} + \frac{1}{n} \sum_{\alpha, k, l, \beta} X_{\alpha k} O_{k\beta} [L^{-1}\Psi]_{\beta\beta} \frac{\partial O'_{\beta l}}{\partial X_{\alpha l}}. \end{aligned}$$

Using (6.2) and (6.3), we obtain

$$\begin{aligned}
&= \sum_{\gamma} \frac{\psi_{\gamma}}{l_{\gamma}} + \frac{1}{n^2} \sum_{\alpha, k, l, \beta, \gamma \neq \beta, \epsilon} X_{\alpha k} O_{k\gamma} \frac{O'_{\gamma l} O_{\epsilon\beta} + O'_{\gamma\epsilon} O_{l\beta}}{l_{\beta} - l_{\gamma}} X_{\alpha\epsilon} [L^{-1}\Psi]_{\beta\beta} O'_{\beta l} \\
&\quad + \frac{2}{n^2} \sum_{\alpha, k, l, \beta, \gamma, \epsilon} X_{\alpha k} O_{k\beta} O'_{\gamma l} X_{\alpha\epsilon} O_{\epsilon\gamma} \frac{\partial [\psi_{\beta}/l_{\beta}]}{\partial l_{\gamma}} O'_{\beta l} \\
&\quad + \frac{1}{n^2} \sum_{\alpha, k, l, \beta, \gamma \neq \beta, \epsilon} X_{\alpha k} O_{k\beta} [L^{-1}\Psi]_{\beta\beta} O_{l\gamma} \frac{O'_{\gamma l} O_{\epsilon\beta} + O'_{\gamma\epsilon} O_{l\beta}}{l_{\beta} - l_{\gamma}} X_{\alpha\epsilon} \\
&= \sum_{\gamma} \frac{\psi_{\gamma}}{l_{\gamma}} + \frac{1}{n} \sum_{\gamma \neq \beta} \frac{l_{\gamma} \psi_{\beta}}{(l_{\beta} - l_{\gamma}) l_{\beta}} + \frac{2}{n} \sum_{\gamma} l_{\gamma} \frac{\partial [\psi_{\gamma}/l_{\gamma}]}{\partial l_{\gamma}} + \frac{1}{n} \sum_{\gamma \neq \beta} \frac{\psi_{\beta}}{l_{\beta} - l_{\gamma}} \\
&= \frac{n-p-1}{n} \sum_{\gamma} \frac{\psi_{\gamma}}{l_{\gamma}} + \frac{2}{n} \sum_{\gamma} \frac{\partial \psi_{\gamma}}{\partial l_{\gamma}} + \frac{1}{n} \sum_{\gamma \neq \beta} \frac{\psi_{\beta} - \psi_{\gamma}}{l_{\beta} - l_{\gamma}}.
\end{aligned}$$

By (6.1), we conclude

$$\begin{aligned}
&\mathbb{E} \left[\left| \text{div}_{\text{vec}(\tilde{X})} \text{vec}(\tilde{X}H) \right| \right] \\
&= \mathbb{E} \left[\left| \sum_{k=1}^p \frac{n-p-1}{n} \frac{\psi_k}{l_k} + \frac{2}{n} \sum_{k=1}^p \frac{\partial \psi_k}{\partial l_k} + \frac{1}{n} \sum_{k \neq b}^p \frac{\psi_k - \psi_b}{l_k - l_b} \right| \right] < \infty.
\end{aligned}$$

Therefore, we can apply Lemma 3 in Chételat and Wells [2012] to $G = \frac{1}{n} O L^{-1} \Psi O'$, which holds for any (p, n) . We obtain that

$$\begin{aligned}
\mathbb{E}[\text{tr}(\Sigma^{-1} O \Psi O')] &= \mathbb{E}[\text{tr}(L^{-1} \Psi) + \text{tr}(X' \nabla_X G')] \\
&= \mathbb{E} \left[\sum_{\gamma} \frac{\psi_{\gamma}}{l_{\gamma}} + \frac{1}{n} \sum_{k, l, \alpha} X'_{k\alpha} \frac{\partial}{\partial X_{\alpha l}} O L^{-1} \Psi O'_{kl} \right].
\end{aligned}$$

But the expression inside the expected value is precisely eq. (6.4), so

$$= \mathbb{E} \left[\sum_{k=1}^p \frac{n-p-1}{n} \frac{\psi_k}{l_k} + \frac{2}{n} \sum_{k=1}^p \frac{\partial \psi_k}{\partial l_k} + \frac{1}{n} \sum_{k \neq b}^p \frac{\psi_k - \psi_b}{l_k - l_b} \right]$$

as desired. \square

Lemma 4. Let $W \sim W_p(n, \Sigma)$ with $n \geq p$, and let $W/n = OLO'$ be the spectral decomposition of the associated sample covariance matrix. Let $\psi_1(L), \dots, \psi_p(L)$ be twice-differentiable functions of the eigenvalues of W/n , and define the associated quantities

$$\psi_k^* = \frac{n-p-1}{n} \frac{\psi_k^2}{l_k} + \frac{4}{n} \psi_k \frac{\partial \psi_k}{\partial l_k} + \frac{2}{n} \psi_k \sum_{b \neq k}^p \frac{\psi_k - \psi_b}{l_k - l_b} \quad \text{for } k = 1, \dots, p$$

with $\Psi^* = \text{diag}(\psi_1, \dots, \psi_p)$. Assume

$$\begin{aligned} \mathbb{E} \left[\sum_{k=1}^p \left| \frac{\psi_k^*}{l_k} \right|^{1+\epsilon} \right] &< \infty \quad \text{and} \\ \mathbb{E} \left[\sum_{k=1}^p \frac{n-p-1}{n} \frac{\psi_k^*}{l_k} + \frac{2}{n} \sum_{k=1}^p \frac{\partial \psi_k^*}{\partial l_k} + \frac{1}{n} \sum_{k \neq b}^p \frac{\psi_k^* - \psi_b^*}{l_k - l_b} \right] &< \infty. \end{aligned}$$

for some $\epsilon > 0$. Then

$$\mathbb{E} \left[\text{tr} \left([\Sigma^{-1} O \Psi O']^2 \right) \right] = \mathbb{E} \left[\sum_{k=1}^p \frac{n-p-1}{n} \frac{\psi_k^*}{l_k} + \frac{2}{n} \sum_{k=1}^p \frac{\partial \psi_k^*}{\partial l_k} + \frac{1}{n} \sum_{k \neq b}^p \frac{\psi_k^* - \psi_b^*}{l_k - l_b} \right].$$

Proof. We use Lemma 3 in [Chételat and Wells \[2012\]](#) again. Decompose $W = X'X$ for some $X \sim N_{n \times p}(0, I_n \otimes \Sigma)$, and define $\tilde{X} = X\Sigma^{-1/2}$ and $H = \frac{1}{n} \Sigma^{1/2} O L^{-1} \Psi O' \Sigma^{-1} O \Psi O' \Sigma^{-1/2}$. Then

$$\begin{aligned} \text{div}_{\text{vec}(\tilde{X})} \text{vec} \left(\tilde{X} H \right) &= \sum_{\alpha, i, j} \frac{\partial}{\partial \tilde{X}_{\alpha i}} \left\{ \tilde{X}_{\alpha j} H_{ji} \right\} = n \sum_i H_{ii} + \sum_{\alpha, j} \tilde{X}_{\alpha j} \frac{\partial H_{ji}}{\partial \tilde{X}_{\alpha i}} \\ (6.5) \end{aligned}$$

$$\begin{aligned} &= \sum_{i, j, \gamma} \Sigma_{ij}^{-1} O_{j\gamma} \frac{\psi_\gamma^2}{l_\gamma} O'_{\gamma i} \\ &\quad + \frac{1}{n} \sum_{\alpha, \beta, i, j, k, l} \tilde{X}_{\alpha j} \Sigma_{\beta i}^{1/2} \Sigma_{jk}^{1/2} \frac{\partial}{\partial X_{\alpha \beta}} \{ O L^{-1} \Psi O' \Sigma^{-1} O \Psi O' \}_{kl} \Sigma_{li}^{-1/2} \\ (6.6) \end{aligned}$$

$$\begin{aligned} &= \sum_{i, j, \gamma} \Sigma_{ij}^{-1} O_{j\gamma} \frac{\psi_\gamma^2}{l_\gamma} O'_{\gamma i} + \frac{1}{n} \sum_{\alpha, k, l} X_{\alpha k} \frac{\partial}{\partial X_{\alpha l}} O L^{-1} \Psi O'_{ki} \Sigma_{ij}^{-1} O \Psi O'_{jl} \\ &= \sum_{i, j, \gamma} \Sigma_{ij}^{-1} O_{j\gamma} \frac{\psi_\gamma^2}{l_\gamma} O'_{\gamma i} + \frac{1}{n} \sum_{i, j, \alpha, k, l, \beta} \Sigma_{ij}^{-1} O \Psi O'_{jl} X_{\alpha k} \frac{\partial O_{k\beta}}{\partial X_{\alpha l}} [L^{-1} \Psi]_{\beta \beta} O'_{\beta i} \\ &\quad + \frac{1}{n} \sum_{i, j, \alpha, k, l, \beta} \Sigma_{ij}^{-1} O \Psi O'_{jl} X_{\alpha k} O_{k\beta} \frac{\partial [L^{-1} \Psi]_{\beta \beta}}{\partial X_{\alpha l}} O'_{\beta i} \\ &\quad + \frac{1}{n} \sum_{i, j, \alpha, k, l, \beta} \Sigma_{ij}^{-1} O \Psi O'_{jl} X_{\alpha k} O_{k\beta} [L^{-1} \Psi]_{\beta \beta} \frac{\partial O_{i\beta}}{\partial X_{\alpha l}} \\ &\quad + \frac{1}{n} \sum_{i, j, \alpha, k, l} X_{\alpha k} O L^{-1} \Psi O'_{ki} \Sigma_{ij}^{-1} \frac{\partial O_{j\beta}}{\partial X_{\alpha l}} \Psi_{\beta \beta} O'_{\beta l} \\ &\quad + \frac{1}{n} \sum_{i, j, \alpha, k, l} X_{\alpha k} O L^{-1} \Psi O'_{ki} \Sigma_{ij}^{-1} O_{j\beta} \frac{\partial \Psi_{\beta \beta}}{\partial X_{\alpha l}} O'_{\beta l} \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{n} \sum_{i,j,\alpha,k,l} X_{\alpha k} O L^{-1} \Psi O'_{ki} \Sigma_{ij}^{-1} O_{j\beta} \Psi_{\beta\beta} \frac{\partial O_{l\beta}}{\partial X_{\alpha l}} \\
& = \sum_{i,j,\gamma} \Sigma_{ij}^{-1} O_{j\gamma} \frac{\psi_\gamma^2}{l_\gamma} O'_{\gamma i} \\
& + \frac{1}{n^2} \sum_{i,j,k,l,\alpha,\beta,\gamma \neq \beta,\epsilon} \Sigma_{ij}^{-1} O \Psi O'_{jl} X_{\alpha k} O_{k\gamma} \frac{O'_{\gamma l} O_{\epsilon\beta} + O'_{\gamma\epsilon} O_{l\beta}}{l_\beta - l_\gamma} X_{\alpha\epsilon} [L^{-1} \Psi]_{\beta\beta} O'_{\beta i} \\
& + \frac{2}{n^2} \sum_{i,j,\alpha,\beta,\gamma,\epsilon} \Sigma_{ij}^{-1} O \Psi O'_{jl} X_{\alpha k} O_{k\beta} O'_{\gamma l} X_{\alpha\epsilon} O_{\epsilon\gamma} \frac{\partial [\psi_\beta / l_\beta]}{\partial l_\gamma} O'_{\beta i} \\
& + \frac{1}{n^2} \sum_{i,j,k,l,\alpha,\beta,\gamma \neq \beta,\epsilon} \Sigma_{ij}^{-1} O \Psi O'_{jl} X_{\alpha k} O_{k\beta} [L^{-1} \Psi]_{\beta\beta} O_{i\gamma} \frac{O'_{\gamma l} O_{\epsilon\beta} + O'_{\gamma\epsilon} O_{l\beta}}{l_\beta - l_\gamma} X_{\alpha\epsilon} \\
& + \frac{1}{n^2} \sum_{i,j,\alpha,\beta,\gamma \neq \beta,\epsilon} X_{\alpha k} O L^{-1} \Psi O'_{ki} \Sigma_{ij}^{-1} O_{j\gamma} \frac{O'_{\gamma l} O_{\epsilon\beta} + O'_{\gamma\epsilon} O_{l\beta}}{l_\beta - l_\gamma} X_{\alpha\epsilon} \Psi_{\beta\beta} O'_{\beta l} \\
& + \frac{2}{n^2} \sum_{i,j,\alpha,\beta,\gamma,\epsilon} X_{\alpha k} O L^{-1} \Psi O'_{ki} \Sigma_{ij}^{-1} O_{j\beta} O'_{\gamma l} X_{\alpha\epsilon} O_{\epsilon\gamma} \frac{\partial \psi_\beta}{\partial l_\gamma} O'_{\beta l} \\
& + \frac{1}{n^2} \sum_{i,j,k,l,\alpha,\beta,\gamma \neq \beta,\epsilon} X_{\alpha k} O L^{-1} \Psi O'_{ki} \Sigma_{ij}^{-1} O_{j\beta} \Psi_{\beta\beta} O_{l\gamma} \frac{O'_{\gamma l} O_{\epsilon\beta} + O'_{\gamma\epsilon} O_{l\beta}}{l_\beta - l_\gamma} X_{\alpha\epsilon} \\
& = \sum_{i,j,\gamma} \Sigma_{ij}^{-1} O_{j\gamma} \frac{\psi_\gamma^2}{l_\gamma} O'_{\gamma i} \\
& + \frac{1}{n} \sum_{i,j,l,\beta,\gamma \neq \beta} \Sigma_{ij}^{-1} O_{j\beta} \frac{l_\gamma \psi_\beta^2}{(l_\beta - l_\gamma) l_\beta} O'_{\beta i} + \frac{2}{n} \sum_{i,j,\gamma} \Sigma_{ij}^{-1} O_{j\gamma} \psi_\gamma l_\gamma \frac{\partial [\psi_\gamma / l_\gamma]}{\partial l_\gamma} O'_{\gamma i} \\
& + \frac{1}{n} \sum_{i,j,\beta,\gamma \neq \beta} \Sigma_{ij}^{-1} O_{j\gamma} \frac{\psi_\beta \psi_\gamma}{l_\beta - l_\gamma} O'_{\gamma i} + \frac{1}{n} \sum_{i,j,\beta,\gamma \neq \beta} \Sigma_{ij}^{-1} O_{j\gamma} \frac{\psi_\gamma \psi_\beta}{l_\beta - l_\gamma} O'_{\gamma i} \\
& + \frac{2}{n} \sum_{i,j,\gamma} \Sigma_{ij}^{-1} O_{j\gamma} \psi_\gamma \frac{\partial \psi_\gamma}{\partial l_\gamma} O'_{\gamma i} + \frac{1}{n} \sum_{i,j,\beta,\gamma \neq \beta,\epsilon} \Sigma_{ij}^{-1} O_{j\beta} \frac{\psi_\beta^2}{l_\beta - l_\gamma} O'_{\beta i} \\
& = \frac{n-p-1}{n} \sum_{i,j,\gamma} \Sigma_{ij}^{-1} O_{j\gamma} \frac{\psi_\gamma^2}{l_\gamma} O'_{\gamma i} + \frac{4}{n} \sum_{i,j,\gamma} \Sigma_{ij}^{-1} O_{j\gamma} \psi_\gamma \frac{\partial \psi_\gamma}{\partial l_\gamma} O'_{\gamma i} \\
(6.7) \quad & + \frac{2}{n} \sum_{i,j,l,\beta,\gamma \neq \beta} \Sigma_{ij}^{-1} O_{j\gamma} \psi_\gamma \frac{(\psi_\gamma - \psi_\beta)}{(l_\gamma - l_\beta)} O'_{\gamma i}
\end{aligned}$$

Thus

$$\begin{aligned}
\mathbb{E} \left[\left| \text{div}_{\text{vec}(\tilde{X})} \text{vec}(\tilde{X}H) \right| \right] &= \frac{1}{n} \mathbb{E} \left[\left| \sum_{i,j=1}^p \Sigma_{ij}^{-1} O \Psi^* O'_{ji} \right| \right] \\
&\leq \frac{1}{n} \mathbb{E} \left[\sum_{k=1}^p \left| \left[L^{1/2} O' \Sigma^{-1} O L^{1/2} \right]_{kk} \left| \frac{\psi_k^*}{l_k} \right| \right| \right] \\
&\leq \frac{1}{n} \sum_{k=1}^p \mathbb{E} \left[\left[L^{1/2} O' \Sigma^{-1} O L^{1/2} \right]_{kk}^{1+\frac{1}{\epsilon}} \right]^{\frac{\epsilon}{1+\epsilon}} \mathbb{E} \left[\left| \frac{\psi_k^*}{l_k} \right|^{1+\epsilon} \right]^{\frac{1}{1+\epsilon}} \\
&\leq \frac{1}{n} \left(\mathbb{E} \left[\sum_{k=1}^p \left[L^{1/2} O' \Sigma^{-1} O L^{1/2} \right]_{kk}^{1+\frac{1}{\epsilon}} \right] \right)^{\frac{\epsilon}{1+\epsilon}} \left(\mathbb{E} \left[\sum_{k=1}^p \left| \frac{\psi_k^*}{l_k} \right|^{1+\epsilon} \right] \right)^{\frac{1}{1+\epsilon}}
\end{aligned}$$

Since

$$\begin{aligned}
\sum_{k=1}^p \left[L^{1/2} O' \Sigma^{-1} O L^{1/2} \right]_{kk}^{1+\frac{1}{\epsilon}} &\leq \left(\sum_{k=1}^p \left| L^{1/2} O' \Sigma^{-1} O L^{1/2} \right|_{kk} \right)^{1+\frac{1}{\epsilon}} \\
&= \text{tr} \left(L^{1/2} O' \Sigma^{-1} O L^{1/2} \right)^{1+\frac{1}{\epsilon}} = \text{tr}(\Sigma^{-1} S)^{1+\frac{1}{\epsilon}} \sim (\chi_{pn}^2)^{1+\frac{1}{\epsilon}}
\end{aligned}$$

we get

$$\begin{aligned}
&\mathbb{E} \left[\left| \text{div}_{\text{vec}(\tilde{X})} \text{vec}(\tilde{X}H) \right| \right] \\
&\leq \frac{2\Gamma \left(1 + \frac{1}{\epsilon} + \frac{np}{2} \right)^{\frac{\epsilon}{1+\epsilon}}}{n\Gamma \left(\frac{np}{2} \right)^{\frac{\epsilon}{1+\epsilon}}} \left(\mathbb{E} \left[\sum_{k=1}^p \left| \frac{\psi_k^*}{l_k} \right|^{1+\epsilon} \right] \right)^{\frac{1}{1+\epsilon}} < \infty
\end{aligned}$$

by assumption of the lemma. Therefore by Lemma 3 in [Chételat and Wells \[2012\]](#),

$$\begin{aligned}
&\mathbb{E} \left[\text{tr} \left([\Sigma^{-1} O \Psi O']^2 \right) \right] = \mathbb{E} \left[\text{tr}(L^{-1} \Psi) + \text{tr}(X' \nabla_X G') \right] \\
&= \mathbb{E} \left[\sum_{i,j,k=1}^p \Sigma_{ij}^{-1} O_{jk} \frac{\psi_k^2}{l_k} O'_{ki} \right. \\
&\quad \left. + \frac{1}{n} \sum_{\alpha=1}^n \sum_{k,l=1}^p X_{\alpha k} \frac{\partial}{\partial X_{\alpha l}} O L^{-1} \Psi O'_{ki} \Sigma_{ij}^{-1} O \Psi O'_{jl} \right] \\
&= \mathbb{E} \left[\sum_{i,j,k=1}^p \Sigma_{ij}^{-1} O_{jk} \psi_k^* O'_{ki} \right] \quad (\text{ by (6.6) }).
\end{aligned}$$

Finally, by Lemma 3, we conclude

$$= \mathbb{E} \left[\sum_{k=1}^p \frac{n-p-1}{n} \frac{\psi_k^*}{l_k} + \frac{2}{n} \sum_{k=1}^p \frac{\partial \psi_k^*}{\partial l_k} + \frac{1}{n} \sum_{k \neq b}^p \frac{\psi_k^* - \psi_b^*}{l_k - l_b} \right]$$

as desired. \square

Lemma 5. *Let $l_1 > \dots > l_p > 0$ be the eigenvalues of a $W_p(n, \Sigma)$ -distributed matrix, for some $\Sigma > 0$. If $n \geq p+1$, then*

- (i) *for any $1 \leq k \leq p$ and $0 \leq m < \frac{n-p-1}{2}$, $\mathbb{E} \left[\frac{1}{|l_k|^m} \right] < \infty$;*
- (ii) *for any $1 \leq k \neq b \leq p$ and $1 \leq m < 2$, $\mathbb{E} \left[\frac{1}{|l_k - l_b|^m} \right] < \infty$;*
- (iii) *for any $1 \leq k \neq b \neq e \leq p$ and $1 \leq m < 2$, $\mathbb{E} \left[\frac{1}{|l_k - l_b|^m |l_k - l_e|^m} \right] < \infty$.*

Proof. First, notice that in (ii), we can take $k < b$ without loss of generality. Then

$$\mathbb{E} \left[\frac{1}{|l_k - l_b|^m} \right] \leq \mathbb{E} \left[\frac{1}{|l_k - l_{k+1}|^m} \right],$$

and it would be enough to show the r.h.s. finite for all $1 \leq k < p$ to show (ii).

Similarly, in (iii), we can take $b < e$ without loss of generality, and there are then three possibilities. Either $k < b < e$, in which case

$$\mathbb{E} \left[\frac{1}{|l_k - l_b|^m |l_k - l_e|^m} \right] \leq \mathbb{E} \left[\frac{1}{|l_k - l_{k+1}|^m |l_{e-1} - l_e|^m} \right],$$

or $b < k < e$ in which case

$$\mathbb{E} \left[\frac{1}{|l_k - l_b|^m |l_k - l_e|^m} \right] \leq \mathbb{E} \left[\frac{1}{|l_b - l_{b+1}|^m |l_{e-1} - l_e|^m} \right]$$

or $b < e < k$ in which case

$$\mathbb{E} \left[\frac{1}{|l_k - l_b|^m |l_k - l_e|^m} \right] \leq \mathbb{E} \left[\frac{1}{|l_b - l_{b+1}|^m |l_e - l_{e+1}|^m} \right]$$

Thus in any case it is enough to show that

$$\mathbb{E} \left[\frac{1}{|l_k - l_{k+1}|^m |l_b - l_{b+1}|^m} \right] < \infty$$

for all $1 \leq k < b < p$ to show (iii).

By Muirhead [1982], Theorem 3.2.18, the joint density of $l_1 > \dots > l_p$ is given by

$$(6.8) \quad f_{l_1, \dots, l_p}(l_1, \dots, l_p) = \frac{\pi^{p^2/2} 2^{-pn} |\Sigma|^{-n/2}}{\Gamma_p(p/2) \Gamma_p(n/2)} \prod_{i=1}^p l_i^{\frac{n-p-1}{2}} \prod_{1 \leq i < j \leq p} (l_i - l_j)$$

$$(6.9) \quad \int_{O(p)} \text{etr} \left(-\frac{1}{2} \Sigma^{-1} H L H' \right) dH \mathbb{1}[l_1 > \dots > l_p > 0]$$

for $L = \text{diag}(l_1, \dots, l_p)$. Define $I_2 = \{(i, j) \mid i < j \wedge (i, j) \neq (k, k+1)\}$ and $I_3 = \{(i, j) \mid i < j \wedge (i, j) \neq (k, k+1), (b, b+1)\}$. The expressions

$$\begin{aligned} P_1(l_1, \dots, l_p) &= \prod_{i \neq k}^p l_i^{n-p-1} \prod_{i < j}^p (l_i - l_j)^2 \\ P_2(l_1, \dots, l_p) &= \prod_{i=1}^p l_i^{n-p-1} \prod_{(i,j) \in I_2} (l_i - l_j)^2 \\ P_3(l_1, \dots, l_p) &= \prod_{i=1}^p l_i^{n-p-1} \prod_{(i,j) \in I_3} (l_i - l_j)^2 \end{aligned}$$

and $K = \frac{\pi^{p^2/2} 2^{-pn} |\Sigma|^{-n/2}}{\Gamma_p(p/2) \Gamma_p(n/2)}$ can then be defined to write

$$\begin{aligned} f_{l_1, \dots, l_p}(l_1, \dots, l_p) &= K l_k^{\frac{n-p-1}{2}} P_1^{1/2}(l_1, \dots, l_p) \\ &\quad \int_{O(p)} \text{etr} \left(-\frac{1}{2} \Sigma^{-1} H L H' \right) dH \mathbb{1}[l_1 > \dots > l_p > 0] \\ &= K |l_k - l_{k+1}| P_2^{1/2}(l_1, \dots, l_p) \\ &\quad \int_{O(p)} \text{etr} \left(-\frac{1}{2} \Sigma^{-1} H L H' \right) dH \mathbb{1}[l_1 > \dots > l_p > 0] \\ &= K |l_k - l_{k+1}| |l_b - l_{b+1}| P_3^{1/2}(l_1, \dots, l_p) \\ &\quad \int_{O(p)} \text{etr} \left(-\frac{1}{2} \Sigma^{-1} H L H' \right) dH \mathbb{1}[l_1 > \dots > l_p > 0]. \end{aligned}$$

The important point is that since $n - p - 1 \geq 0$, P_1 , P_2 and P_3 are polynomials in l_1, \dots, l_p . Define $x = l_{k+1} - l_k$, $y = l_{k+1} - l_b$ and $z = l_b - l_{b+1}$, so that $l_k = l_{b+1} + z + y + x$, $l_{k+1} = l_{b+1} + z + y$ and $l_b = l_{b+1} + z$. (It might happen that $k+1=b$, something which should be kept in mind.) Then

$$\begin{aligned} &P_2(l_1, \dots, l_{k+1} + x, \dots, l_{k+1}, \dots, l_p), \\ &P_3(l_1, \dots, l_{b+1} + z + x, \dots, l_{b+1} + z, \dots, l_b) \quad \text{if } k+1 = b, \\ &P_3(l_1, \dots, l_{b+1} + z + y + x, \dots, l_{b+1} + z + y, \dots, l_{b+1} + z, \dots, l_b) \quad \text{if } k+1 \neq b, \end{aligned}$$

must still be polynomials, in $\{l_i\} \setminus \{l_k\} \cup \{x\}$, $\{l_i\} \setminus \{l_k, l_b\} \cup \{x, z\}$ and $\{l_i\} \setminus \{l_k, l_{k+1}, l_b\} \cup \{x, y, z\}$ respectively. Therefore, for some finite degrees D_1, \dots, D_4 one can write

$$\begin{aligned} P_1(l_1, \dots, l_p) &= \sum_{\substack{d_1 + \dots + d_p \\ \leq D_1}} A_{d_1, \dots, d_p}^1 l_1^{d_1} \dots l_p^{d_p}, \\ P_2(l_1, \dots, l_{k+1} + x, \dots, l_{k+1}, \dots, l_p) &= \sum_{\substack{d_1 + \dots + d_p \\ \leq D_2}} A_{d_1, \dots, d_p}^2 l_1^{d_1} \dots x^{d_k} \dots l_p^{d_p}, \\ P_3(l_1, \dots, l_{b+1} + z + x, \dots, l_{b+1} + z, \dots, l_b) & \end{aligned}$$

$$\begin{aligned}
&= \sum_{\substack{d_1+\dots+d_p \\ \leq D_3}} A_{d_1,\dots,d_p}^3 l_1^{d_1} \dots x^{d_k} \dots z^{d_b} \dots l_p^{d_p}, & \text{if } k+1 = b, \\
P_3(l_1, \dots, l_{b+1} + z + y + x, \dots, l_{b+1} + z + y, \dots, l_{b+1} + z, \dots, l_b) \\
&= \sum_{\substack{d_1+\dots+d_p \\ \leq D_4}} A_{d_1,\dots,d_p}^4 l_1^{d_1} \dots x^{d_k} \dots y^{d_{k+1}} \dots z^{d_b} \dots l_p^{d_p}, & \text{if } k+1 \neq b.
\end{aligned}$$

for coefficients $A_{d_1,\dots,d_p}^1, \dots, A_{d_1,\dots,d_p}^4 \in \mathbb{R}$.

If we denote the greatest eigenvalue of Σ by λ_{\max} , then $\Sigma^{-1} \geq \lambda_{\max}^{-1} I$ so that

$$\int_{O(p)} \text{etr} \left(-\frac{1}{2} \Sigma^{-1} H L H' \right) dH \leq \exp \left(-\frac{1}{2\lambda_{\max}} \sum_{i=1}^p l_i \right)$$

for any $l_1, \dots, l_p \geq 0$.

Now, for (i), we can use (6.9) and $\mathbb{1}[l_1 > \dots > l_p > 0] \leq \prod_{i=1}^p \mathbb{1}[l_i > 0]$ to find

$$\begin{aligned}
\mathbb{E} \left[\frac{1}{|l_k|^m} \right] &\leq K \int_{\mathbb{R}^p} \frac{1}{l_k^{m-\frac{n-p-1}{2}}} P_1^{1/2}(l_1, \dots, l_p) \exp \left(-\frac{1}{2\lambda_{\max}} \sum_{i=1}^p l_i \right) \\
&\quad \mathbb{1}[l_1 > \dots > l_p > 0] dl_1 \dots dl_p \\
&\leq K \sum_{\substack{d_1+\dots+d_p \\ \leq D_1}} \sqrt{|A_{d_1,\dots,d_p}^1|} \int_0^\infty l_1^{d_1/2} e^{-l_1/2\lambda_{\max}} dl_1 \dots \\
&\quad \int_0^\infty l_k^{d_k/2-m+\frac{n-p-1}{2}} e^{-l_k/2\lambda_{\max}} dl_k \dots \int_0^\infty l_p^{d_p/2} e^{-l_p/2\lambda_{\max}} dl_p.
\end{aligned}$$

Notice that for $i \neq k$, $\int_0^\infty l_i^{d_i/2} e^{-l_i/2\lambda_{\max}} dl_i < \infty$ for any $d_i \geq 0$, and $\int_0^\infty l_k^{d_k/2-m+\frac{n-p-1}{2}} e^{-l_k/2\lambda_{\max}} dl_k < \infty$ for all $d_k \geq 0$ iff $0 \leq m < \frac{n-p-1}{2}$. Thus $\mathbb{E}[1/|l_k|^m] < \infty$, as desired.

For (ii), we proceed similarly, but though a change of variables $(l_k, l_{k+1}) \rightarrow (l_{k+1} + x, l_k)$. Then, using

$$\mathbb{1} \left[l_1 > \dots > l_{k+1} + x > \dots > l_{k+1} > \dots > l_p > 0 \right] \leq \mathbb{1}[x > 0] \prod_{i \neq k} \mathbb{1}[l_i > 0],$$

we obtain

$$\begin{aligned}
\mathbb{E} \left[\frac{1}{|l_k - l_{k+1}|^m} \right] &\leq K \int_{\mathbb{R}^p} \frac{1}{|l_k - l_{k+1}|^{m-1}} P_2^{1/2}(l_1, \dots, l_p) \\
&\quad \exp \left(-\frac{1}{2\lambda_{\max}} \sum_{i=1}^p l_i \right) \mathbb{1}[l_1 > \dots > l_p > 0] dl_1 \dots dl_p
\end{aligned}$$

$$\begin{aligned}
&\leq K \sum_{\substack{d_1+\dots+d_p \\ \leq D_2}} \sqrt{|A_{d_1,\dots,d_p}^2|} \int_0^\infty l_1^{d_1/2} e^{-l_1/2\lambda_{\max}} dl_1 \dots \\
&\quad \int_0^\infty x^{d_k/2-m+1} e^{-x/2\lambda_{\max}} dx \dots \int_0^\infty l_{k+1}^{d_{k+1}/2} e^{-l_{k+1}/\lambda_{\max}} dl_{k+1} \dots \\
&\quad \int_0^\infty l_p^{d_p/2} e^{-l_p/2\lambda_{\max}} dl_p.
\end{aligned}$$

Then again, for $i \neq k$ and any $d_i \geq 0$, the respective integrals are finite; and to have $\int_0^\infty x^{d_k/2-m+1} e^{-x/2\lambda_{\max}} dx < \infty$ for all $d_k \geq 0$ requires $m < 2$. In such a case, we end up with $E[1/|l_k - l_{k+1}|^m] < \infty$, as desired.

For (iii), we must consider separately the cases $k+1 = b$ and $k+1 \neq b$. In the first case, one can take the change of variables $(l_k, l_b, l_{b+1}) \rightarrow (l_{b+1} + x + z, l_{b+1} + z, l_{b+1})$. Using that

$$\begin{aligned}
&\mathbb{1} \left[l_1 > \dots > l_{b+1} + x + z > \dots > l_{b+1} + z > \dots > l_{b+1} > \dots > l_p > 0 \right] \\
&\leq \mathbb{1}[x > 0] \mathbb{1}[z > 0] \prod_{i \neq k, b}^p \mathbb{1}[l_i > 0],
\end{aligned}$$

we then obtain

$$\begin{aligned}
&E \left[\frac{1}{|l_k - l_b|^m |l_b - l_{b+1}|^m} \right] \leq K \int_{\mathbb{R}^p} \frac{1}{|l_k - l_b|^{m-1} |l_b - l_{b+1}|^{m-1}} \\
&\quad P_3^{1/2}(l_1, \dots, l_p) \exp \left(-\frac{1}{2\lambda_{\max}} \sum_{i=1}^p l_i \right) \mathbb{1}[l_1 > \dots > l_p > 0] dl_1 \dots dl_p \\
&\leq K \sum_{\substack{d_1+\dots+d_p \\ \leq D_3}} \sqrt{|A_{d_1,\dots,d_p}^3|} \int_0^\infty l_1^{d_1/2} e^{-l_1/2\lambda_{\max}} dl_1 \dots \\
&\quad \int_0^\infty x^{d_k/2-m+1} e^{-x/2\lambda_{\max}} dx \dots \int_0^\infty z^{d_b/2-m+1} e^{-z/\lambda_{\max}} dz \dots \\
&\quad \int_0^\infty l_{b+1}^{d_{b+1}/2} e^{-3l_{b+1}/2\lambda_{\max}} dl_{b+1} \dots \int_0^\infty l_p^{d_p/2} e^{-l_p/2\lambda_{\max}} dl_p.
\end{aligned}$$

Again, all the integrals converge as long as $m < 2$, in which case we have $E[1/|l_k - l_b|^m |l_b - l_{b+1}|^m] < \infty$, as desired.

Finally, for (iii) with $k+1 \neq b$, one can take the change of variables $(l_k, l_{k+1}, l_b, l_{b+1}) \rightarrow (l_{b+1} + x + y + z, l_{b+1} + y + z, l_{b+1} + z, l_{b+1})$. Then using

that

$$\begin{aligned} & \mathbb{1} \left[l_1 > \dots > l_{b+1} + x + y + z > \dots > l_{b+1} + y + z > \dots > l_{b+1} + z \right. \\ & \quad \left. > \dots > l_{b+1} > \dots > l_p > 0 \right] \\ & \leq \mathbb{1}[x > 0] \mathbb{1}[y > 0] \mathbb{1}[z > 0] \prod_{i \neq k, k+1, b}^p \mathbb{1}[l_i > 0], \end{aligned}$$

we obtain

$$\begin{aligned} \mathbb{E} \left[\frac{1}{|l_k - l_{k+1}|^m |l_b - l_{b+1}|^m} \right] & \leq K \int_{\mathbb{R}^p} \frac{1}{|l_k - l_{k+1}|^{m-1} |l_b - l_{b+1}|^{m-1}} \\ & \quad P_4^{1/2}(l_1, \dots, l_p) \exp \left(-\frac{1}{2\lambda_{\max}} \sum_{i=1}^p l_i \right) \mathbb{1}[l_1 > \dots > l_p > 0] dl_1 \dots dl_p \\ & \leq K \sum_{\substack{d_1 + \dots + d_p \\ \leq D_4}} \sqrt{|A_{d_1, \dots, d_p}^4|} \int_0^\infty l_1^{d_1/2} e^{-l_1/2\lambda_{\max}} dl_1 \dots \\ & \quad \int_0^\infty x^{d_k/2-m+1} e^{-x/2\lambda_{\max}} dx \dots \int_0^\infty y^{d_k/2-m+1} e^{-y/\lambda_{\max}} dy \dots \\ & \quad \int_0^\infty z^{d_b/2-m+1} e^{-3z/2\lambda_{\max}} dz \dots \int_0^\infty l_{b+1}^{d_{b+1}/2} e^{-2l_{b+1}/\lambda_{\max}} dl_{b+1} \dots \\ & \quad \int_0^\infty l_p^{d_p/2} e^{-l_p/2\lambda_{\max}} dl_p. \end{aligned}$$

All the integrals converge as long as $m < 2$, in which case we have $\mathbb{E}[1/|l_k - l_{k+1}|^m |l_b - l_{b+1}|^m] < \infty$, as desired. \square

Proof of Theorem 1. By independence, it is clear that

$$\begin{aligned} \mathbb{E} \left[L_H(\hat{\Sigma}, \Sigma) \right] & = \mathbb{E}_{\hat{\rho}} \left[\mathbb{E}_S \left[L_H(\hat{\Sigma}, \Sigma) \mid \hat{\rho} \right] \right] \\ & = \mathbb{E}_{\hat{\rho}} \left[\mathbb{E}_S \left[L_H(\hat{\Sigma}, \Sigma) \right] \right], \end{aligned}$$

so we can treat $\hat{\rho}$ as a constant throughout the calculations, without loss of generality. Define the auxiliary terms $\psi_k = \hat{\gamma}_k + \hat{\sigma}^2$ and

$$\psi_k^* = \frac{n-p-1}{n} \frac{\psi_k^2}{l_k} + 4 \frac{\psi_k}{n} \frac{\partial \psi_k}{\partial l_k} + 2 \frac{\psi_k}{n} \sum_{b \neq k}^p \frac{\psi_k - \psi_b}{l_k - l_b}$$

for all $1 \leq k \leq p$, and consider:

$$\begin{aligned}
R_1 &= \sum_{k=1}^p \frac{n-p-1}{n} \frac{\psi_k^*}{l_k} + \frac{2}{n} \sum_{k=1}^p \frac{\partial \psi_k^*}{\partial l_k} + \frac{1}{n} \sum_{k \neq b}^p \frac{\psi_k^* - \psi_b^*}{l_k - l_b} \\
&= \left\{ \frac{(n-p-1)^2}{n^2} \sum_{k=1}^p \frac{\psi_k^2}{l_k^2} + 4 \frac{n-p-1}{n^2} \sum_{k=1}^p \frac{\psi_k}{l_k} \frac{\partial \psi_k}{\partial l_k} \right. \\
&\quad \left. + 2 \frac{n-p-1}{n^2} \sum_{k \neq b=1}^p \frac{\psi_k}{l_k} \frac{\psi_k - \psi_b}{l_k - l_b} \right\} + \left\{ 4 \frac{n-p-1}{n^2} \sum_{k=1}^p \frac{\psi_k}{l_k} \frac{\partial \psi_k}{\partial l_k} \right. \\
&\quad - 2 \frac{n-p-1}{n^2} \sum_{k=1}^p \frac{\psi_k^2}{l_k^2} + \frac{8}{n^2} \sum_{k=1}^p \left(\frac{\partial \psi_k}{\partial l_k} \right)^2 + \frac{8}{n^2} \sum_{k=1}^p \psi_k \frac{\partial^2 \psi_k}{\partial l_k^2} \\
&\quad + \frac{4}{n^2} \sum_{k \neq b=1}^p \frac{\partial \psi_k}{\partial l_k} \frac{\psi_k - \psi_b}{l_k - l_b} + \frac{4}{n^2} \sum_{k \neq b=1}^p \psi_k \frac{\frac{\partial \psi_k}{\partial l_k} - \frac{\partial \psi_b}{\partial l_b}}{l_k - l_b} \\
&\quad \left. - \frac{2}{n^2} \sum_{k \neq b=1}^p \left(\frac{\psi_k - \psi_b}{l_k - l_b} \right)^2 \right\} + \left\{ \left(2 \frac{n-p-1}{n^2} \sum_{k \neq b=1}^p \frac{\psi_k}{l_k} \frac{\psi_k - \psi_b}{l_k - l_b} \right. \right. \\
&\quad \left. - \frac{n-p-1}{n^2} \sum_{k \neq b=1}^p \frac{\psi_k}{l_k} \frac{\psi_b}{l_b} \right) + \left(\frac{4}{n^2} \sum_{k \neq b=1}^p \psi_k \frac{\frac{\partial \psi_k}{\partial l_k} - \frac{\partial \psi_b}{\partial l_b}}{l_k - l_b} \right. \\
&\quad \left. + \frac{4}{n^2} \sum_{k \neq b=1}^p \frac{\partial \psi_k}{\partial l_k} \frac{\psi_k - \psi_b}{l_k - l_b} \right) + \left(\frac{2}{n^2} \sum_{k \neq b \neq e=1}^p \frac{\psi_k}{l_k - l_b} \left(\frac{\psi_k - \psi_e}{l_k - l_e} - \frac{\psi_b - \psi_e}{l_b - l_e} \right) \right. \\
&\quad \left. + \frac{2}{n^2} \sum_{k \neq b \neq e=1}^p \frac{\psi_k - \psi_b}{l_k - l_b} \frac{\psi_k - \psi_e}{l_k - l_e} + \frac{2}{n^2} \sum_{k \neq b=1}^p \left(\frac{\psi_k - \psi_b}{l_k - l_b} \right)^2 \right) \Big\} \\
&= \frac{(n-p-1)(n-p-2)}{n^2} \sum_{k=1}^p \frac{\psi_k^2}{l_k^2} - \frac{(n-p-1)}{n^2} \left(\sum_{k=1}^p \frac{\psi_k}{l_k} \right)^2 \\
&\quad + \frac{8}{n^2} \sum_{k=1}^p \left(\frac{\partial \psi_k}{\partial l_k} \right)^2 + \frac{8}{n^2} \sum_{k=1}^p \psi_k \frac{\partial^2 \psi_k}{\partial l_k^2} + 8 \frac{n-p-1}{n^2} \sum_{k=1}^p \frac{\psi_k}{l_k} \frac{\partial \psi_k}{\partial l_k} \\
&\quad + 4 \frac{n-p-1}{n^2} \sum_{k \neq b=1}^p \frac{\psi_k}{l_k} \frac{\psi_k - \psi_b}{l_k - l_b} + \frac{8}{n^2} \sum_{k \neq b=1}^p \frac{\partial \psi_k}{\partial l_k} \frac{\psi_k - \psi_b}{l_k - l_b} \\
&\quad + \frac{4}{n^2} \sum_{k \neq b=1}^p \psi_k \frac{\frac{\partial \psi_k}{\partial l_k} - \frac{\partial \psi_b}{\partial l_b}}{l_k - l_b} + \frac{4}{n^2} \sum_{k \neq b=1}^p \psi_k \frac{\frac{\partial \psi_k}{\partial l_k} - \frac{\partial \psi_b}{\partial l_b}}{l_k - l_b} \\
&\quad + \frac{2}{n^2} \sum_{k \neq b \neq e=1}^p \frac{\psi_k}{l_k - l_b} \left(\frac{\psi_k - \psi_e}{l_k - l_e} - \frac{\psi_b - \psi_e}{l_b - l_e} \right)
\end{aligned}$$

$$+ \frac{2}{n^2} \sum_{k \neq b \neq e=1}^p \frac{\psi_k - \psi_b}{l_k - l_b} \frac{\psi_k - \psi_e}{l_k - l_e}.$$

Now, by Hölder's inequality, we find:

$$\begin{aligned} \mathbb{E}[|R_1|] &\leq \frac{|n-p-1||n-p-2|}{n^2} \sum_{k=1}^p \mathbb{E}\left[\left|\frac{\psi_k}{l_k}\right|^2\right] \\ &\quad + \frac{|n-p-1|}{n^2} \left(\sum_{k=1}^p \mathbb{E}\left[\left|\frac{\psi_k}{l_k}\right|^2\right]^{\frac{1}{2}}\right)^2 + \frac{8}{n^2} \sum_{k=1}^p \mathbb{E}\left[\left|\frac{\partial \psi_k}{\partial l_k}\right|^2\right] \\ &\quad + \frac{8}{n^2} \sum_{k=1}^p \mathbb{E}\left[\left|\psi_k \frac{\partial^2 \psi_k}{\partial l_k^2}\right|\right] + 8 \frac{|n-p-1|}{n^2} \sum_{k=1}^p \mathbb{E}\left[\left|\frac{\psi_k}{l_k}\right|^2\right]^{\frac{1}{2}} \mathbb{E}\left[\left|\frac{\partial \psi_k}{\partial l_k}\right|^2\right]^{\frac{1}{2}} \\ &\quad + 4 \frac{|n-p-1|}{n^2} \sum_{k \neq b=1}^p \mathbb{E}\left[\left|\frac{\psi_k}{l_k}\right|^{4.5}\right]^{\frac{1}{4.5}} \\ &\quad \quad \left(\mathbb{E}\left[|\psi_k|^{4.5}\right]^{\frac{1}{4.5}} + \mathbb{E}\left[|\psi_b|^{4.5}\right]^{\frac{1}{4.5}}\right) \mathbb{E}\left[\frac{1}{|l_k - l_b|^{1.8}}\right]^{\frac{1}{1.8}} \\ &\quad + \frac{8}{n^2} \sum_{k \neq b=1}^p \left(\mathbb{E}\left[\left|\frac{\partial \psi_k}{\partial l_k}\right|^{4.5}\right]^{\frac{1}{4.5}} \mathbb{E}\left[|\psi_k|^{4.5}\right]^{\frac{1}{4.5}}\right. \\ &\quad \quad \left.+ \mathbb{E}\left[\left|\frac{\partial \psi_k}{\partial l_k}\right|^{4.5}\right]^{\frac{1}{4.5}} \mathbb{E}\left[|\psi_b|^{4.5}\right]^{\frac{1}{4.5}}\right) \mathbb{E}\left[\frac{1}{|l_k - l_b|^{1.8}}\right]^{\frac{1}{1.8}} \\ &\quad + \frac{4}{n^2} \sum_{k \neq b=1}^p \mathbb{E}\left[|\psi_k|^{4.5}\right]^{\frac{1}{4.5}} \left(\mathbb{E}\left[\left|\frac{\partial \psi_k}{\partial l_k}\right|^{4.5}\right]^{\frac{1}{4.5}}\right. \\ &\quad \quad \left.+ \mathbb{E}\left[\left|\frac{\partial \psi_b}{\partial l_b}\right|^{4.5}\right]^{\frac{1}{4.5}}\right) \mathbb{E}\left[\frac{1}{|l_k - l_b|^{1.8}}\right]^{\frac{1}{1.8}} \\ &\quad + \frac{4}{n^2} \sum_{k \neq b=1}^p \mathbb{E}\left[|\psi_k|^{4.5}\right]^{\frac{1}{4.5}} \left(\mathbb{E}\left[\left|\frac{\partial \psi_k}{\partial l_k}\right|^{4.5}\right]^{\frac{1}{4.5}}\right. \\ &\quad \quad \left.+ \mathbb{E}\left[\left|\frac{\partial \psi_b}{\partial l_k}\right|^{4.5}\right]^{\frac{1}{4.5}}\right) \mathbb{E}\left[\frac{1}{|l_k - l_b|^{1.8}}\right]^{\frac{1}{1.8}} \\ &\quad + \frac{2}{n^2} \sum_{k \neq b \neq e=1}^p \left(\mathbb{E}\left[|\psi_k|^{4.5}\right]^{\frac{1}{4.5}} + \mathbb{E}\left[|\psi_b|^{4.5}\right]^{\frac{1}{4.5}}\right) \end{aligned}$$

$$\begin{aligned}
& \left(\mathbb{E} \left[|\psi_k|^{4.5} \right]^{\frac{1}{4.5}} + \mathbb{E} \left[|\psi_e|^{4.5} \right]^{\frac{1}{4.5}} \right) \mathbb{E} \left[\left| \frac{1}{(l_k - l_b)(l_k - l_e)} \right|^{1.8} \right]^{\frac{1}{1.8}} \\
& + \frac{2}{n^2} \sum_{k \neq b \neq e=1}^p \left(\mathbb{E} \left[|\psi_k|^{4.5} \right]^{\frac{1}{4.5}} + \mathbb{E} \left[|\psi_b|^{4.5} \right]^{\frac{1}{4.5}} \right) \\
& \left(\mathbb{E} \left[|\psi_k|^{4.5} \right]^{\frac{1}{4.5}} + \mathbb{E} \left[|\psi_e|^{4.5} \right]^{\frac{1}{4.5}} \right) \mathbb{E} \left[\left| \frac{1}{(l_k - l_b)(l_k - l_e)} \right|^{1.8} \right]^{\frac{1}{1.8}}.
\end{aligned}$$

Similarly, consider

$$R_2 = \frac{n-p-1}{n} \sum_{k=1}^p \frac{\psi_k}{l_k} + \frac{2}{n} \sum_{k=1}^p \frac{\partial \psi_k}{\partial l_k} + \frac{1}{n} \sum_{k \neq b=1}^p \frac{\psi_k - \psi_b}{l_k - l_b},$$

so that

$$\begin{aligned}
\mathbb{E} \left[|R_2| \right] & \leq \frac{|n-p-1|}{n} \sum_{k=1}^p \mathbb{E} \left[\left| \frac{\psi_k}{l_k} \right| \right] + \frac{2}{n} \sum_{k=1}^p \mathbb{E} \left[\left| \frac{\partial \psi_k}{\partial l_k} \right| \right] \\
& + \frac{1}{n} \sum_{k \neq b=1}^p \left(\mathbb{E} \left[|\psi_k|^{2.25} \right]^{\frac{1}{2.25}} + \mathbb{E} \left[|\psi_b|^{2.25} \right]^{\frac{1}{2.25}} \right) \mathbb{E} \left[\frac{1}{|l_k - l_b|^{1.8}} \right]^{\frac{1}{1.8}},
\end{aligned}$$

Moreover, for any $\epsilon > 0$,

$$\begin{aligned}
\mathbb{E} \left[\sum_{k=1}^p \left| \frac{\psi_k^*}{l_k} \right| \right] & \leq \frac{|n-p-1|}{n} \sum_{k=1}^p \mathbb{E} \left[\left| \frac{\psi_k}{l_k} \right|^{2(1+\epsilon)} \right]^{\frac{1}{1+\epsilon}} + \frac{4}{n} \sum_{k=1}^p \mathbb{E} \left[\left| \frac{\psi_k}{l_k} \right|^{2(1+\epsilon)} \right] \\
& \cdot \mathbb{E} \left[\left| \frac{\partial \psi_k}{\partial l_k} \right|^{2(1+\epsilon)} \right]^{\frac{1}{2(1+\epsilon)}} + \frac{2}{n} \sum_{k \neq b}^p \mathbb{E} \left[\frac{1}{|l_k - l_b|^{1.8(1+\epsilon)}} \right]^{\frac{1}{1.8(1+\epsilon)}} \\
& \cdot \mathbb{E} \left[\left| \frac{\psi_k}{l_k} \right|^{4.5(1+\epsilon)} \right]^{\frac{1}{4.5(1+\epsilon)}} \left(\mathbb{E} \left[|\psi_k|^{4.5(1+\epsilon)} \right]^{\frac{1}{4.5(1+\epsilon)}} + \mathbb{E} \left[|\psi_b|^{4.5(1+\epsilon)} \right]^{\frac{1}{4.5(1+\epsilon)}} \right).
\end{aligned}$$

Note that for any $1 \leq k \leq p$ and $\epsilon > 0$,

$$\mathbb{E} \left[|\psi_k|^{4.5} \right]^{\frac{1}{4.5}} \leq \mathbb{E} \left[\left| \frac{\psi_k}{l_k} \right|^{9(1+\epsilon)} \right]^{\frac{1}{9(1+\epsilon)}} \mathbb{E} \left[|l_k|^{\frac{9\epsilon}{1+\epsilon}} \right]^{\frac{1+\epsilon}{9\epsilon}},$$

and for any $m > 1$, $\mathbb{E}[|l_k|^m]^2 \leq \mathbb{E}[\text{tr}(S^{2m})] \leq \text{tr}(\Sigma)^{2m} \mathbb{E}[\text{tr}(S\Sigma^{-1})^{2m}] = \text{tr}(\Sigma)^{2m} \mathbb{E}[(\chi_{np}^2)^{2m}] < \infty$. Now consider that, for any $m > 1$,

$$\begin{aligned} |\psi_k|^m &\leq 2^{m-1}(|\hat{\gamma}_k|^m + |\hat{\sigma}^2|^m) \\ \left| \frac{\partial \psi_k}{\partial l_k} \right|^m &\leq 2^{m-1} \left(\left| \frac{\partial \hat{\gamma}_k}{\partial l_k} \right|^m + \left| \frac{\partial \hat{\sigma}^2}{\partial l_k} \right|^m \right) \\ \left| \psi_k \frac{\partial^2 \psi_k}{\partial l_k^2} \right|^m &= |\hat{\gamma}_k + \hat{\sigma}^2| \left| \frac{\partial^2 \hat{\gamma}_k}{\partial l_k^2} + \frac{\partial^2 \hat{\sigma}^2}{\partial l_k^2} \right|^m. \end{aligned}$$

Therefore, since $\hat{\Gamma}$ satisfies the weak regularity conditions and $\hat{\Sigma} \in V_p(\hat{\Gamma})$, we obtain $\mathbb{E}[|R_1|] < \infty$, $\mathbb{E}[|R_2|] < \infty$ and $\mathbb{E}\left[\sum_{k=1}^p \left| \frac{\psi_k^*}{l_k} \right|\right] < \infty$.

Therefore, all the regularity conditions of Lemmas 3 and 4 are satisfied, and we have for $\Psi = \text{diag}(\psi_1, \dots, \psi_p)$

$$\begin{aligned} \mathbb{E}[L(\hat{\Sigma}, \Sigma)] &= \frac{1}{p} \mathbb{E}[\text{tr}([\Sigma^{-1}O\Psi O']^2) - 2\text{tr}(\Sigma^{-1}O\Psi O') + p] \\ &= \frac{1}{p} \mathbb{E}[R_1 - 2R_2 + p] \\ &= \mathbb{E}\left[\frac{(n-p-1)(n-p-2)}{n^2p} \sum_{k=1}^p \frac{\psi_k^2}{l_k^2} - \frac{(n-p-1)}{n^2p} \left(\sum_{k=1}^p \frac{\psi_k}{l_k} \right)^2 \right. \\ &\quad + \frac{8}{n^2p} \sum_{k=1}^p \left(\frac{\partial \psi_k}{\partial l_k} \right)^2 + \frac{8}{n^2p} \sum_{k=1}^p \psi_k \frac{\partial^2 \psi_k}{\partial l_k^2} + 8 \frac{n-p-1}{n^2p} \sum_{k=1}^p \frac{\psi_k}{l_k} \frac{\partial \psi_k}{\partial l_k} \\ &\quad + 4 \frac{n-p-1}{n^2p} \sum_{k \neq b=1}^p \frac{\psi_k}{l_k} \frac{\psi_k - \psi_b}{l_k - l_b} + \frac{8}{n^2p} \sum_{k \neq b=1}^p \frac{\partial \psi_k}{\partial l_k} \frac{\psi_k - \psi_b}{l_k - l_b} \\ &\quad + \frac{4}{n^2p} \sum_{k \neq b=1}^p \psi_k^n \frac{\frac{\partial \psi_k}{\partial l_k} - \frac{\partial \psi_b}{\partial l_b}}{l_k - l_b} + \frac{4}{n^2p} \sum_{k \neq b=1}^p \psi_k^n \frac{\frac{\partial \psi_k}{\partial l_k} - \frac{\partial \psi_b}{\partial l_b}}{l_k - l_b} \\ &\quad + \frac{2}{n^2p} \sum_{k \neq b \neq e=1}^p \frac{\psi_k}{l_k - l_b} \left(\frac{\psi_k - \psi_e}{l_k - l_e} - \frac{\psi_b - \psi_e}{l_b - l_e} \right) \\ &\quad + \frac{2}{n^2p} \sum_{k \neq b \neq e=1}^p \frac{\psi_k - \psi_b}{l_k - l_b} \frac{\psi_k - \psi_e}{l_k - l_e} - 2 \frac{n-p-1}{np} \sum_{k=1}^p \frac{\psi_k}{l_k} \\ &\quad \left. - \frac{4}{np} \sum_{k=1}^p \frac{\partial \psi_k}{\partial l_k} - \frac{2}{np} \sum_{k \neq b=1}^p \frac{\psi_k - \psi_b}{l_k - l_b} + 1 \right]. \end{aligned}$$

We can now collect the terms of order 1 and $1/p$, defining

$$F(\hat{\Sigma}) = \frac{(n-p-1)(n-p-2)}{n^2p} \sum_{k=1}^p \frac{\psi_k^2}{l_k^2} - \frac{n-p-1}{n^2p} \left(\sum_{k=1}^p \frac{\psi_k}{l_k} \right)^2$$

$$\begin{aligned}
& + 4 \frac{n-p-1}{n^2 p} \sum_{k \neq b=1}^p \frac{\psi_k}{l_k} \frac{\psi_k - \psi_b}{l_k - l_b} + \frac{2}{n^2 p} \sum_{k \neq b \neq e=1}^p \frac{\psi_k - \psi_b}{l_k - l_b} \frac{\psi_k - \psi_e}{l_k - l_e} \\
& + \frac{2}{n^2 p} \sum_{k \neq b \neq e=1}^p \frac{\psi_k}{l_k - l_b} \left(\frac{\psi_k - \psi_e}{l_k - l_e} - \frac{\psi_b - \psi_e}{l_b - l_e} \right) \\
& - 2 \frac{n-p-1}{n p} \sum_{k=1}^p \frac{\psi_k}{l_k} - \frac{2}{n p} \sum_{k \neq b=1}^p \frac{\psi_k - \psi_b}{l_k - l_b} + 1
\end{aligned}$$

and

$$\begin{aligned}
G(\hat{\Sigma}) &= \frac{8}{n^2 p} \sum_{k=1}^p \left(\frac{\partial \psi_k}{\partial l_k} \right)^2 + \frac{8}{n^2 p} \sum_{k=1}^p \psi_k \frac{\partial^2 \psi_k}{\partial l_k^2} + 8 \frac{n-p-1}{n^2 p} \sum_{k=1}^p \frac{\psi_k}{l_k} \frac{\partial \psi_k}{\partial l_k} \\
& + \frac{8}{n^2 p} \sum_{k \neq b=1}^p \frac{\partial \psi_k}{\partial l_k} \frac{\psi_k - \psi_b}{l_k - l_b} + \frac{4}{n^2 p} \sum_{k \neq b=1}^p \psi_k \frac{\frac{\partial \psi_k}{\partial l_k} - \frac{\partial \psi_b}{\partial l_b}}{l_k - l_b} \\
& + \frac{4}{n^2 p} \sum_{k \neq b=1}^p \psi_k \frac{\frac{\partial \psi_k}{\partial l_k} - \frac{\partial \psi_b}{\partial l_b}}{l_k - l_b} - \frac{4}{n p} \sum_{k=1}^p \frac{\partial \psi_k}{\partial l_k}
\end{aligned}$$

so that $\mathbb{E}[L(\hat{\Sigma}, \Sigma)] = \mathbb{E}[F(\hat{\Sigma}) + G(\hat{\Sigma})]$, with $\mathbb{E}[\|F(\hat{\Sigma}) + G(\hat{\Sigma})\|] \leq \mathbb{E}[\|R_1\|] + 2\mathbb{E}[\|R_2\|] + p < \infty$, as desired. Plugging in $\psi_k = \hat{\gamma}_k + \hat{\sigma}^2$ yields, after a bit of algebra:

$$\begin{aligned}
F(\hat{\Gamma} + \hat{\sigma}^2 I) &= \frac{(n-p-1)(n-p-2)}{n^2 p} \sum_{k=1}^p \frac{\hat{\gamma}_k^2}{l_k^2} \\
& + 2 \frac{(n-p-1)(n-p-2)}{n^2 p} \sum_{k=1}^p \frac{\hat{\gamma}_k \hat{\sigma}^2}{l_k^2} + \frac{(n-p-1)(n-p-2)}{n^2 p} \sum_{c=1}^p \frac{\hat{\sigma}_c^4}{l_c^2} \\
& - \frac{n-p-1}{n^2 p} \left(\sum_{k=1}^p \frac{\hat{\gamma}_k}{l_k} \right)^2 - 2 \frac{n-p-1}{n^2 p} \sum_{c=1}^p \frac{\hat{\sigma}^2}{l_c} \sum_{k=1}^p \frac{\hat{\gamma}_k}{l_k} \\
& - \frac{n-p-1}{n^2 p} \left(\sum_{c=1}^p \frac{\hat{\sigma}^2}{l_c} \right)^2 + 4 \frac{n-p-1}{n^2 p} \sum_{k \neq b}^p \frac{\hat{\gamma}_k}{l_k} \frac{\hat{\gamma}_k - \hat{\gamma}_b}{l_k - l_b} \\
& + 4 \frac{n-p-1}{n^2 p} \sum_{k=1}^p \sum_{c=\rho+1}^p \frac{\hat{\gamma}_k}{l_k} \frac{\hat{\gamma}_k}{l_k - l_c} + 4 \frac{n-p-1}{n^2 p} \sum_{k=1}^p \sum_{c=\rho+1}^p \frac{\hat{\sigma}^2}{l_c} \frac{\hat{\gamma}_k}{l_k - l_c} \\
& + \frac{2}{n^2 p} \sum_{k \neq b \neq e=1}^p \frac{\hat{\gamma}_k - \hat{\gamma}_b}{l_k - l_b} \frac{\hat{\gamma}_k - \hat{\gamma}_e}{l_k - l_e} + \frac{2}{n^2 p} \sum_{k \neq b=1}^p \sum_{c=\rho+1}^p \frac{\hat{\gamma}_k}{l_k - l_c} \frac{\hat{\gamma}_b}{l_b - l_c} \\
& - \frac{6}{n^2 p} \sum_{k \neq b}^p \sum_{c=\rho+1}^p \frac{\hat{\gamma}_k + \hat{\sigma}^2}{l_k - l_c} \frac{\hat{\gamma}_b}{l_b - l_c} + \frac{6}{n^2 p} \sum_{k=1}^p \sum_{c \neq d}^p \frac{\hat{\gamma}_k + \hat{\sigma}^2}{l_k - l_c} \frac{\hat{\gamma}_k}{l_k - l_d}
\end{aligned}$$

$$\begin{aligned}
& -\frac{2}{n^2 p} \sum_{k=1}^{\rho} \sum_{c \neq d}^p \frac{\hat{\gamma}_k}{l_k - l_c} \frac{\hat{\gamma}_k}{l_k - l_d} + \frac{2}{n^2 p} \sum_{k \neq b \neq e=1}^{\rho} \frac{\hat{\gamma}_k}{l_k - l_b} \left(\frac{\hat{\gamma}_k - \hat{\gamma}_e}{l_k - l_e} - \frac{\hat{\gamma}_b - \hat{\gamma}_e}{l_b - l_e} \right) \\
& + \frac{4}{n^2 p} \sum_{k \neq b=1}^p \sum_{c=\rho+1}^p \frac{\hat{\gamma}_k - \hat{\gamma}_b}{l_k - l_b} \frac{\hat{\gamma}_k}{l_k - l_c} + \frac{6}{n^2 p} \sum_{k \neq b}^{\rho} \sum_{c=\rho+1}^p \frac{\hat{\gamma}_k - \hat{\gamma}_b}{l_k - l_b} \frac{\hat{\gamma}_k + \hat{\sigma}^2}{l_k - l_c} \\
& - 2 \frac{n-p-1}{np} \sum_{k=1}^{\rho} \frac{\hat{\gamma}_k}{l_k} - 2 \frac{n-p-1}{np} \sum_{c=1}^p \frac{\hat{\sigma}^2}{l_c} - \frac{2}{np} \sum_{k \neq b=1}^{\rho} \frac{\hat{\gamma}_k - \hat{\gamma}_b}{l_k - l_b} \\
(6.10) \quad & - \frac{4}{np} \sum_{k=1}^{\rho} \sum_{c=\rho+1}^p \frac{\hat{\gamma}_k}{l_k - l_c} + 1
\end{aligned}$$

and

$$\begin{aligned}
G(\hat{\Gamma} + \hat{\sigma}^2 I) &= \frac{8}{n^2 p} \sum_{k=1}^{\rho} \left(\frac{\partial \hat{\gamma}_k}{\partial l_k} \right)^2 + \frac{16}{n^2 p} \sum_{k=1}^{\rho} \frac{\partial \hat{\gamma}_k}{\partial l_k} \frac{\partial \hat{\sigma}^2}{\partial l_k} + \frac{8}{n^2 p} \sum_{k=1}^p \left(\frac{\partial \hat{\sigma}^2}{\partial l_k} \right)^2 \\
& + \frac{8}{n^2 p} \sum_{k=1}^{\rho} \hat{\gamma}_k \frac{\partial^2 \hat{\gamma}_k}{\partial l_k^2} + \frac{8}{n^2 p} \sum_{k=1}^p \hat{\sigma}^2 \frac{\partial^2 \hat{\sigma}^2}{\partial l_k^2} + 8 \frac{n-p-1}{n^2 p} \sum_{k=1}^{\rho} \frac{\hat{\gamma}_k}{l_k} \frac{\partial \hat{\gamma}_k}{\partial l_k} \\
& + 8 \frac{n-p-1}{n^2 p} \sum_{k=1}^{\rho} \frac{\hat{\gamma}_k}{l_k} \frac{\partial \hat{\sigma}^2}{\partial l_k} + 8 \frac{n-p-1}{n^2 p} \sum_{k=1}^p \frac{\hat{\sigma}^2}{l_k} \frac{\partial \hat{\gamma}_k}{\partial l_k} \\
& + 8 \frac{n-p-1}{n^2 p} \sum_{k=1}^p \frac{\hat{\sigma}^2}{l_k} \frac{\partial \hat{\sigma}^2}{\partial l_k} + \frac{8}{n^2 p} \sum_{k \neq b=1}^{\rho} \frac{\partial \hat{\gamma}_k}{\partial l_k} \frac{\hat{\gamma}_k - \hat{\gamma}_b}{l_k - l_b} \\
& + \frac{8}{n^2 p} \sum_{k=1}^{\rho} \sum_{c=\rho+1}^p \frac{\partial \hat{\gamma}_k}{\partial l_k} \frac{\hat{\gamma}_k}{l_k - l_c} + \frac{8}{n^2 p} \sum_{k=1}^{\rho} \sum_{c=\rho+1}^p \frac{\partial \hat{\sigma}^2}{\partial l_c} \frac{\hat{\gamma}_k}{l_k - l_c} \\
& + \frac{4}{n^2 p} \sum_{k \neq b=1}^{\rho} \hat{\gamma}_k \frac{\frac{\partial \hat{\gamma}_k}{\partial l_k} - \frac{\partial \hat{\gamma}_b}{\partial l_b}}{l_k - l_b} + \frac{4}{n^2 p} \sum_{k=1}^{\rho} \sum_{c=\rho+1}^p \hat{\gamma}_k \frac{\frac{\partial \hat{\gamma}_k}{\partial l_k} - \frac{\partial \hat{\sigma}^2}{\partial l_c}}{l_k - l_c} \\
& + \frac{4}{n^2 p} \sum_{k=1}^{\rho} \sum_{c=\rho+1}^p \hat{\sigma}^2 \frac{\frac{\partial \hat{\gamma}_k}{\partial l_k} - \frac{\partial \hat{\sigma}^2}{\partial l_c}}{l_k - l_c} + \frac{4}{n^2 p} \sum_{c \neq d=\rho+1}^p \hat{\sigma}^2 \frac{\frac{\partial \hat{\sigma}^2}{\partial l_c} - \frac{\partial \hat{\sigma}^2}{\partial l_d}}{l_c - l_d} \\
& + \frac{4}{n^2 p} \sum_{k \neq b=1}^{\rho} \hat{\gamma}_k \frac{\frac{\partial \hat{\gamma}_k}{\partial l_k} - \frac{\partial \hat{\gamma}_b}{\partial l_b}}{l_k - l_b} + \frac{4}{n^2 p} \sum_{k=1}^{\rho} \sum_{c=\rho+1}^p \hat{\gamma}_k \frac{\frac{\partial \hat{\gamma}_k}{\partial l_k} - \frac{\partial \hat{\sigma}^2}{\partial l_c}}{l_k - l_c} \\
& + \frac{4}{n^2 p} \sum_{k=1}^{\rho} \sum_{c=\rho+1}^p \hat{\sigma}^2 \frac{\frac{\partial \hat{\gamma}_k}{\partial l_c} - \frac{\partial \hat{\sigma}^2}{\partial l_c}}{l_k - l_c} + \frac{4}{n^2 p} \sum_{c \neq d=\rho+1}^p \hat{\sigma}^2 \frac{\frac{\partial \hat{\sigma}^2}{\partial l_c} - \frac{\partial \hat{\sigma}^2}{\partial l_d}}{l_c - l_d} \\
(6.11) \quad & - \frac{4}{np} \sum_{k=1}^{\rho} \frac{\partial \hat{\gamma}_k}{\partial l_k} - \frac{4}{np} \sum_{k=1}^p \frac{\partial \hat{\sigma}^2}{\partial l_k}.
\end{aligned}$$

For the second part of the theorem, we see that

$$\begin{aligned}
\mathbb{E} \left[\left| F(\hat{\Sigma}) \right| \right] &\leq 1 + \frac{|n-p-1||n-p-2|}{n^2} \mathbb{E} \left[\sup_{p \in \mathbb{N}^*} \max_{1 \leq k \leq p} \left| \frac{\psi_k}{l_k} \right|^2 \right] \\
&+ \frac{|n-p-1|p}{n^2} \mathbb{E} \left[\sup_{p \in \mathbb{N}^*} \max_{1 \leq k \leq p} \left| \frac{\psi_k}{l_k} \right|^2 \right] \\
&+ 4 \frac{|n-p-1|(p-1)}{n^2} \mathbb{E} \left[\sup_{p \in \mathbb{N}^*} \max_{1 \leq k \leq p} \left| \frac{\psi_k}{l_k} \right|^2 \right]^{\frac{1}{2}} \mathbb{E} \left[\sup_{p \in \mathbb{N}^*} \max_{1 \leq k \neq b \leq p} \left| \frac{\psi_k - \psi_b}{l_k - l_b} \right|^2 \right]^{\frac{1}{2}} \\
&+ \frac{2(p-1)(p-2)}{n^2} \mathbb{E} \left[\sup_{p \in \mathbb{N}^*} \max_{1 \leq k \neq b \leq p} \left| \frac{\psi_k - \psi_b}{l_k - l_b} \right|^2 \right] \\
&+ \frac{2(p-1)(p-2)}{n^2} \mathbb{E} \left[\sup_{p \in \mathbb{N}^*} \max_{1 \leq k \leq p} \left| \frac{\psi_k}{l_k} \right|^2 \right]^{\frac{1}{2}} \\
&\quad \cdot \mathbb{E} \left[\sup_{p \in \mathbb{N}^*} \max_{1 \leq k \neq b \neq e \leq p} \left| \frac{l_k}{l_k - l_b} \left(\frac{\psi_k - \psi_e}{l_k - l_e} - \frac{\psi_b - \psi_e}{l_b - l_e} \right) \right|^2 \right]^{\frac{1}{2}} \\
&- 2 \frac{|n-p-1|}{n} \mathbb{E} \left[\sup_{p \in \mathbb{N}^*} \max_{1 \leq k \leq p} \left| \frac{\psi_k}{l_k} \right| \right] - \frac{2(p-1)}{n} \mathbb{E} \left[\sup_{p \in \mathbb{N}^*} \max_{1 \leq k \neq b \leq p} \left| \frac{\psi_k - \psi_b}{l_k - l_b} \right| \right]
\end{aligned}$$

and

$$\begin{aligned}
p \mathbb{E} \left[\left| G(\hat{\Sigma}) \right| \right] &\leq \frac{8p}{n^2} \mathbb{E} \left[\sup_{p \in \mathbb{N}^*} \max_{1 \leq k \leq p} \left| \frac{\partial \psi_k}{\partial l_k} \right|^2 \right] + \frac{8p}{n^2} \mathbb{E} \left[\sup_{p \in \mathbb{N}^*} \max_{1 \leq k \leq p} \left| \psi_k \frac{\partial^2 \psi_k}{\partial l_k^2} \right| \right] \\
&+ 8 \frac{|n-p-1|p}{n} \mathbb{E} \left[\sup_{p \in \mathbb{N}^*} \max_{1 \leq k \leq p} \left| \frac{\psi_k}{l_k} \right|^2 \right]^{\frac{1}{2}} \mathbb{E} \left[\sup_{p \in \mathbb{N}^*} \max_{1 \leq k \leq p} \left| \frac{\partial \psi_k}{\partial l_k} \right|^2 \right]^{\frac{1}{2}} \\
&+ \frac{8(p-1)p}{n^2} \mathbb{E} \left[\sup_{p \in \mathbb{N}^*} \max_{1 \leq k \leq p} \left| \frac{\partial \psi_k}{\partial l_k} \right|^2 \right]^{\frac{1}{2}} \mathbb{E} \left[\sup_{p \in \mathbb{N}^*} \max_{1 \leq k \neq b \leq p} \left| \frac{\psi_k - \psi_b}{l_k - l_b} \right|^2 \right]^{\frac{1}{2}} \\
&+ \frac{4(p-1)p}{n^2} \mathbb{E} \left[\sup_{p \in \mathbb{N}^*} \max_{1 \leq k \leq p} \left| \frac{\psi_k}{l_k} \right|^2 \right]^{\frac{1}{2}} \mathbb{E} \left[\sup_{p \in \mathbb{N}^*} \max_{\substack{1 \leq k \\ \neq b \leq p}} \left| \frac{l_k}{l_k - l_b} \left(\frac{\partial \psi_k}{\partial l_k} - \frac{\partial \psi_b}{\partial l_b} \right) \right|^2 \right]^{\frac{1}{2}} \\
&+ \frac{4(p-1)p}{n^2} \mathbb{E} \left[\sup_{p \in \mathbb{N}^*} \max_{1 \leq k \leq p} \left| \frac{\psi_k}{l_k} \right|^2 \right]^{\frac{1}{2}} \mathbb{E} \left[\sup_{p \in \mathbb{N}^*} \max_{\substack{1 \leq k \\ \neq b \leq p}} \left| \frac{l_k}{l_k - l_b} \left(\frac{\partial \psi_k}{\partial l_k} - \frac{\partial \psi_b}{\partial l_k} \right) \right|^2 \right]^{\frac{1}{2}} \\
&- \frac{4p}{n} \mathbb{E} \left[\sup_{p \in \mathbb{N}^*} \max_{1 \leq k \leq p} \left| \frac{\partial \psi_k}{\partial l_k} \right| \right].
\end{aligned}$$

Again, one can proceed like in the weak case to see that if $\hat{\Gamma}$ satisfies its strong regularity conditions and $\hat{\Sigma} \in \tilde{V}_p(\hat{\Gamma})$, we get $\lim_{n \rightarrow \infty} \mathbb{E}[|F(\hat{\Sigma})|] < \infty$ and $\lim_{n \rightarrow \infty} p \mathbb{E}[|G(\hat{\Sigma})|] < \infty$ as $\lim_{n \rightarrow \infty} \frac{p_n}{n} \in (0, 1)$, as desired. \square

Proof of Proposition 1. First note any element of $C_c^\infty(H_p; \mathbb{R})$, the space of smooth, compactly supported functions from H_p to \mathbb{R} , satisfies the weak regularity conditions of Definition 2 for any weak $\hat{\Gamma}$. Now, if $\tilde{\Sigma}$ is a minimum over $V_p(\hat{\Gamma})$, then for any $\eta \in C_c^\infty(H_+^p; \mathbb{R})$ and any $t \in \mathbb{R}$, $\tilde{\sigma}^2 + t\eta$ satisfies the $\hat{\Gamma}$ -weak regularity conditions too and $\epsilon \rightarrow \mathbb{E}[F(\hat{\Gamma} + [\tilde{\sigma}^2 + t\eta]I)]$ is smooth over \mathbb{R} with a minimum at $t = 0$. But we find that the first variation satisfies

$$\begin{aligned}
\frac{\partial}{\partial t} \mathbb{E}[F(\hat{\Gamma} + [\tilde{\sigma}^2 + t\eta]I)] \Big|_{t=0} &= \mathbb{E} \left[\eta \cdot \left(2 \frac{(n-p-1)(n-p-2)}{n^2 p} \sum_{k=1}^{\hat{\rho}} \frac{\hat{\gamma}_k}{l_k^2} \right. \right. \\
&\quad + 2 \frac{(n-p-1)(n-p-2)}{n^2 p} \sum_{c=1}^p \frac{\tilde{\sigma}^2}{l_c^2} - 2 \frac{n-p-1}{n^2 p} \sum_{c=1}^p \frac{1}{l_c} \sum_{k=1}^{\hat{\rho}} \frac{\hat{\gamma}_k}{l_k} \\
&\quad - 2 \frac{n-p-1}{n^2 p} \sum_{c=1}^p \frac{1}{l_c} \sum_{c=1}^p \frac{\tilde{\sigma}^2}{l_c} + 4 \frac{n-p-1}{n^2 p} \sum_{k=1}^{\hat{\rho}} \sum_{c=\hat{\rho}+1}^p \frac{1}{l_c} \frac{\hat{\gamma}_k}{l_k - l_c} \\
&\quad - \frac{6}{n^2 p} \sum_{k \neq b}^{\hat{\rho}} \sum_{c=\hat{\rho}+1}^p \frac{1}{l_k - l_c} \frac{\hat{\gamma}_b}{l_b - l_c} + \frac{6}{n^2 p} \sum_{k=1}^{\hat{\rho}} \sum_{c \neq d=\hat{\rho}+1}^p \frac{1}{l_k - l_c} \frac{\hat{\gamma}_k}{l_k - l_d} \\
&\quad \left. + \frac{6}{n^2 p} \sum_{k \neq b}^{\hat{\rho}} \sum_{c=\hat{\rho}+1}^p \frac{\hat{\gamma}_k - \hat{\gamma}_b}{l_k - l_b} \frac{1}{l_k - l_c} - 2 \frac{n-p-1}{np} \sum_{c=1}^p \frac{1}{l_c} \right) \Big] \\
&= \mathbb{E}[\eta \cdot F_1[l, \hat{\rho}, \hat{\gamma}, \tilde{\sigma}^2]] \\
(6.12) \quad &= \int_{H_p} \eta(l_1, \dots, l_p) F_1[l, \hat{\rho}, \hat{\gamma}, \tilde{\sigma}^2] \cdot f_{l_1, \dots, l_p}(l_1, \dots, l_p) \prod_{i=1}^p dl_i,
\end{aligned}$$

where $f_{l_1, \dots, l_p}(l_1, \dots, l_p)$ stands for the p.d.f. of $l_1 > \dots > l_p$. Now, if this equals zero for all $\eta \in C_c^\infty(H_+^p; \mathbb{R})$, by the fundamental lemma of calculus of variations (see, say, [Giaquinta and Hildebrandt \[1996\]](#) ch. 2.2) we obtain $F_1[l, \hat{\rho}, \hat{\gamma}, \tilde{\sigma}^2] \cdot f_{l_1, \dots, l_p}(l_1, \dots, l_p) \equiv 0$, that is, $F_1[l, \hat{\rho}, \hat{\gamma}, \tilde{\sigma}^2] \equiv 0$. This implies $\tilde{\sigma}^2 = A/B$.

For the second statement, notice that by construction, the space of $\hat{\Gamma}$ -weak noise estimators is convex; let $\hat{\sigma}^2$ be some arbitrary element. Define $H : [0, 1] \rightarrow \mathbb{R}$ to be the smooth function

$$H(t) = \mathbb{E}[F(\hat{\Gamma} + [\tilde{\sigma}^2 + t(\hat{\sigma}^2 - \tilde{\sigma}^2)]I)].$$

Notice that, for F_1 as in eq. (6.12),

$$H'(0) = \mathbb{E}[(\hat{\sigma}^2 - \tilde{\sigma}^2) \cdot F_1[l, \hat{\rho}, \hat{\gamma}, \tilde{\sigma}^2]] = 0,$$

since $\tilde{\sigma}^2 = A/B$. Moreover,

$$\begin{aligned} H''(t) &= \frac{\partial^2}{\partial t^2} \mathbb{E}\left[F(\hat{\Gamma}_r + [\tilde{\sigma}^2 + t(\hat{\sigma}^2 - \tilde{\sigma}^2)]I)\right] \\ (6.13) \quad &= 2 \frac{n-p-1}{n^2 p} \mathbb{E}\left[(\hat{\sigma}^2 - \tilde{\sigma}^2)^2 \left((n-p-2) \sum_{c=1}^p \frac{1}{l_c^2} - \left(\sum_{c=1}^p \frac{1}{l_c}\right)^2\right)\right] \\ &\geq 2 \frac{(n-p-1)(n-2p-2)}{n^2 p^2} \mathbb{E}\left[(\hat{\sigma}^2 - \tilde{\sigma}^2)^2 \left(\sum_{c=1}^p \frac{1}{l_c}\right)^2\right] \quad \left(\begin{array}{l} \text{Jensen's} \\ \text{inequality} \end{array}\right) \\ &\geq 0, \end{aligned}$$

for $n \geq 2p + 2$. Therefore, by integration by parts

$$\begin{aligned} &\mathbb{E}\left[F(\hat{\Gamma} + \hat{\sigma}^2 I)\right] - \mathbb{E}\left[F(\hat{\Gamma} + \tilde{\sigma}^2 I)\right] \\ &= H(1) - H(0) = \int_0^1 (1-t) H''(t) dt \geq 0. \end{aligned}$$

Since this is true for any $\hat{\Gamma}$ -weak noise estimator $\hat{\sigma}^2$, we conclude that $\tilde{\Sigma}$ is a minimum over $V_p(\hat{\Gamma})$, as desired. \square

6.2. Proofs for Section 3.

Proof of Lemma 1. To simplify notation in what follows, define $c_{\pm} = [1 \pm \sqrt{c}]^2$. In the proof of Theorem 2.3 in Nadler [2008], p. 2807, it is remarked that for $\sigma^2 = 1$ and $\rho = 1$, the empirical distribution of l_2, \dots, l_p converges a.s. to a Marčenko-Pastur distribution with parameter c . That is, for the truncated empirical spectral measure $d\mu_p = \frac{1}{p-\rho} \sum_{c=\rho+1}^p d\delta_{l_i}$ (where the δ are Dirac measures) we have weak convergence $d\mu_p \Rightarrow d\mu_{\text{MP}(c)}$ a.s. where

$$d\mu_{\text{MP}(c)} = \frac{\sqrt{(c_+ - t)(t - c_-)}}{2\pi c t} \mathbb{1}\left[c_- \leq t \leq c_+\right] dt.$$

As noted by the author, the argument carries on for $\rho \neq 1$, and if $\sigma \neq 1$ we can apply the argument to $l_{\rho+1}/\sigma^2, \dots, l_p/\sigma^2$ to obtain $d\mu_p \Rightarrow d\mu_{\sigma^2 \text{MP}(c)}$ a.s., where

$$d\mu_{\sigma^2 \text{MP}(c)} = \frac{\sqrt{(\sigma^2 c_+ - t)(t - \sigma^2 c_-)}}{2\pi c \sigma^2 t} \mathbb{1}\left[\sigma^2 c_- \leq t \leq \sigma^2 c_+\right] dt.$$

Part (i) Applying the results of Baik and Silverstein [2006], Theorem 1.1 to l_k/σ^2 and $l_{\rho+1}/\sigma^2$ we obtain:

$$l_k \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \frac{(\gamma_k + \sigma^2)(\gamma_k + c\sigma^2)}{\gamma_k}, \quad l_{\rho+1} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} c_+\sigma^2.$$

We will write $\bar{l}_k = (\gamma_k + \sigma^2)(\gamma_k + c\sigma^2)/\gamma_k$ to simplify notation. Let the underlying sample space be denoted Ω . Since $\gamma_\rho > \sqrt{c}\sigma^2$, we have $\bar{l}_k - c_+\sigma^2 = M$ for some $M > 0$. Therefore, for almost all $\omega \in \Omega$, there exists an $N_1(\omega)$ such that $\forall n > N(\omega)$, $l_k^p(\omega) - l_p^p(\omega) > \dots > l_k^p(\omega) - l_{\rho+1}^p(\omega) > M/2$ and $l_p^p(\omega) < \dots < l_{\rho+1}^p(\omega) < c_+\sigma^2 + M/2$. Moreover, for any $\epsilon > 0$, there must be an $N_2(\omega)$ such that for all $n > N_2(\omega, \epsilon)$, $|\bar{l}_k - l_k^p(\omega)| < \epsilon$. Notice that we can write, for any $n > N_1(\omega) \vee N_2(\omega, \epsilon)$,

$$\begin{aligned} \frac{1}{p-\rho} \sum_{c=\rho+1}^p \frac{l_c^p(\omega)}{l_k^p(\omega) - l_c^p(\omega)} &= \frac{1}{p-\rho} \sum_{c=\rho+1}^p \frac{[\bar{l}_k - l_k^p(\omega)]l_c^p(\omega)}{[l_k^p(\omega) - l_c^p(\omega)][\bar{l}_k - l_c^p(\omega)]} \\ &\quad + \int_{(0, c_+\sigma^2 + \frac{M}{2})} \frac{t}{\bar{l}_k - t} d\mu_p(t, \omega), \end{aligned}$$

and

$$(6.14) \quad \left| \frac{1}{p-\rho} \sum_{c=\rho+1}^p \frac{[\bar{l}_k - l_k^p(\omega)]l_c^p(\omega)}{[l_k^p(\omega) - l_c^p(\omega)][\bar{l}_k - l_c^p(\omega)]} \right| < \left[\frac{4}{M^2} c_+\sigma^2 + \frac{2}{M} \right] \epsilon.$$

But $0 < t/(\bar{l}_k - t) < 1 + 2c_+\sigma^2/M$ on $t \in (0, c_+\sigma^2 + \frac{M}{2})$, and it is certainly continuous. Therefore, by the portmanteau theorem of weak convergence of measures,

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_{(0, c_+\sigma^2 + \frac{M}{2})} \frac{t}{\bar{l}_k - t} d\mu_p(t, \omega) &= \int_{(0, c_+\sigma^2 + \frac{M}{2})} \frac{t}{\bar{l}_k - t} d\mu_{\sigma^2 \text{MP}(c)}(t) \\ &= \int_{c_-\sigma^2}^{c_+\sigma^2} \frac{t}{\bar{l}_k - t} \frac{\sqrt{(\sigma^2 c_+ - t)(t - \sigma^2 c_-)}}{2\pi c\sigma^2 t} dt \\ (6.15) \quad &= \frac{2\sigma^2}{\bar{l}_k - (c+1)\sigma^2 + \sqrt{[\bar{l}_k - (c+1)\sigma^2]^2 - 4c\sigma^4}}. \end{aligned}$$

But by definition, $\bar{l}_k = (\gamma_k + \sigma^2)(\gamma_k + c\sigma^2)/\gamma_k$ which can be rewritten as a quadratic equation in γ_k ,

$$\gamma_k^2 - [(\bar{l}_k - (c+1)\sigma^2)]\gamma_k + c\sigma^4 = 0.$$

The roots are

$$\frac{1}{2}[\bar{l}_k - (c+1)\sigma^2] \pm \frac{1}{2}\sqrt{[\bar{l}_k - (c+1)\sigma^2]^2 - 4c\sigma^4},$$

and notice that twice the negative root satisfies

$$\begin{aligned}
& [\bar{l}_k - (c+1)\sigma^2] - \sqrt{[\bar{l}_k - (c+1)\sigma^2]^2 - 4c\sigma^4} \\
&= [\bar{l}_k - (c+1)\sigma^2] \\
&\quad - \sqrt{([\bar{l}_k - (c+1)\sigma^2] - 2\sqrt{c}\sigma^2)([\bar{l}_k - (c+1)\sigma^2] + 2\sqrt{c}\sigma^2)} \\
&\leq [\bar{l}_k - (c+1)\sigma^2] - [\bar{l}_k - (c+1)\sigma^2] + 2\sqrt{c}\sigma^2 \\
&= 2\sqrt{c}\sigma^2.
\end{aligned}$$

Therefore, γ_k cannot equal the negative root, because it would imply $\gamma_k \leq \sqrt{c}\sigma^2$, a contradiction. So γ_k equals the positive root, which, plugged in eq. (6.15), yields

$$\lim_{n \rightarrow \infty} \int_{(0, c_+ \sigma^2 + \frac{M}{2})} \frac{t}{\bar{l}_k - t} d\mu_p(t, \omega) = \frac{\sigma^2}{\gamma_k}.$$

Hence, for some $N_3(\omega, \epsilon)$, we have for all $n > N_3(\omega, \epsilon)$

$$\left| \int_{(0, c_+ \sigma^2 + \frac{M}{2})} \frac{t}{\bar{l}_k - t} d\mu_p(t, \omega) - \frac{\sigma^2}{\gamma_k} \right| < \epsilon.$$

Therefore, from eq. (6.14), we obtain that for $n > N_1(\omega) \vee N_2(\omega, \epsilon) \vee N_3(\omega, \epsilon)$

$$\left| \frac{1}{p - \rho} \sum_{c=\rho+1}^p \frac{l_c^p(\omega)}{\bar{l}_k^p(\omega) - l_c^p(\omega)} - \frac{\sigma^2}{\gamma_k} \right| < \left[\frac{4}{M^2} c_+ \sigma^2 + \frac{2}{M} + 1 \right] \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we conclude that for almost all $\omega \in \Omega$,

$$\lim_{n \rightarrow \infty} \frac{1}{p - \rho} \sum_{c=\rho+1}^p \frac{l_c^p(\omega)}{\bar{l}_k^p(\omega) - l_c^p(\omega)} = \frac{\sigma^2}{\gamma_k},$$

as desired.

Part (ii) The proof is similar in spirit to the previous one, but simpler. Applying the results of Baik and Silverstein [2006], Theorem 1.1 to l_p/σ^2 we obtain:

$$l_p \xrightarrow[n \rightarrow \infty]{\text{a.s.}} c_- \sigma^2.$$

Therefore, for almost all $\omega \in \Omega$, there is a $N(\omega)$ such that $\forall n > N(\omega)$, $l_{\rho+1}^p(\omega) > \dots > l_p^p(\omega) > c_- \sigma^2/2$. Hence, for $n > N(\omega)$,

$$\frac{1}{p - \rho} \sum_{c=\rho+1}^p \frac{1}{\bar{l}_c^{pm}(\omega)} = \int_{\left(\frac{c_- \sigma^2}{2}, \infty\right)} \frac{1}{t^m} d\mu_p(t, \omega),$$

Certainly, $0 < 1/t^m < \left(\frac{2}{c_- \sigma^2}\right)^m$ for $t \in \left(\frac{c_- \sigma^2}{2}, \infty\right)$, and $1/t^m$ is continuous there. Thus, by the portmanteau theorem of weak convergence of measures,

we have for almost all ω

$$\begin{aligned}
\lim_{n \rightarrow \infty} \frac{1}{p - \rho} \sum_{c=\rho+1}^p \frac{1}{l_c^{pm}(\omega)} &= \int_{\left(\frac{c-\sigma^2}{2}, \infty\right)} \frac{1}{t^m} d\mu_{\sigma^2 \text{MP}(c)}(t) \\
&= \int_{c-\sigma^2}^{c+\sigma^2} \frac{1}{t^m} \frac{\sqrt{(\sigma^2 c_+ - t)(t - \sigma^2 c_-)}}{2\pi c \sigma^2 t} dt \\
&= \frac{1}{(1-c)^{2m-1}} \frac{1}{\sigma^{2m}},
\end{aligned}$$

which concludes the proof. \square

Proof of Theorem 2. Recall the definition of $\tilde{\sigma}^2$ as A/B from proposition 1.

Part (i) It follows easily from the Lemma 1 that

$$(4.16) \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \frac{1}{\sigma^2} \quad \text{and} \quad B \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \frac{1}{\sigma^4}.$$

The result then follows immediately.

Part (ii) Denote by $\bar{\gamma}_k$ the a.s. finite limit $\lim_{n \rightarrow \infty} \hat{\gamma}_k$. Start by writing $A =$

$$\frac{n-p-1}{np} \sum_{c=1}^p \frac{1}{l_c} + E_n. \text{ We find}$$

$$\begin{aligned}
&n(\tilde{\sigma}^2 - \sigma^2) \\
&= n\sigma^2 \left[\frac{\frac{n-p-1}{np} \sum_{c=1}^p \frac{\sigma^2}{l_c}}{\frac{(n-p-1)(n-p-2)}{n^2 p} \sum_{c=1}^p \frac{\sigma^4}{l_c^2} - \frac{(n-p-1)p}{n^2} \left(\frac{1}{p} \sum_{c=1}^p \frac{\sigma^2}{l_c} \right)^2} - 1 \right] + \frac{nE_n}{B} \\
&= \frac{\sigma^2}{\sigma^4 B} \left[n \left(\frac{n-p-1}{np} \sum_{c=1}^p \frac{\sigma^2}{l_c} - 1 \right) \right. \\
&\quad \left. + n \left(\frac{1}{1 - \frac{p-\hat{\rho}}{n}} - \frac{(n-p-1)(n-p-2)}{n^2 p} \sum_{c=1}^p \frac{\sigma^4}{l_c^2} \right) \right. \\
&\quad \left. + n \left(\frac{(n-p-1)p}{n^2} \left(\frac{1}{p} \sum_{c=1}^p \frac{\sigma^2}{l_c} \right)^2 - \frac{\frac{p-\hat{\rho}}{n}}{1 - \frac{p-\hat{\rho}}{n}} \right) \right] + \frac{nE_n}{B}.
\end{aligned} \tag{6.17}$$

First consider nE_n . We know the asymptotic behavior of all terms except

$\frac{3}{np} \sum_{k=1}^{\hat{\rho}} \sum_{c \neq d=r+1}^p \frac{1}{l_k - l_c} \frac{\hat{\gamma}_k}{l_k - l_d}$, which we can crudely bound as

$$0 < \frac{3}{np} \sum_{k=1}^{\hat{\rho}} \sum_{c \neq d=\hat{\rho}+1}^p \frac{1}{l_k - l_c} \frac{\hat{\gamma}_k}{l_k - l_d} < \frac{3}{np} \sum_{k=1}^{\hat{\rho}} \hat{\gamma}_k \left(\sum_{c=\hat{\rho}+1}^p \frac{1}{l_k - l_c} \right)^2.$$

Therefore, we obtain that

$$\begin{aligned} & \frac{(1-c)^2}{c} \sum_{k=1}^{\hat{\rho}} \frac{\gamma_k^2 \bar{\gamma}_k}{(\gamma_k + \sigma^2)^2 (\gamma_k + c\sigma^2)^2} + \frac{1}{\sigma^2} \sum_{k=1}^{\hat{\rho}} \frac{\gamma_k \bar{\gamma}_k}{(\gamma_k + \sigma^2)(\gamma_k + c\sigma^2)} \\ & \quad - 2 \frac{1}{\sigma^2} \sum_{k=1}^{\hat{\rho}} \frac{\gamma_k \bar{\gamma}_k}{(\gamma_k + c\sigma^2)^2} \\ & \quad \geq \lim_{n \rightarrow \infty} nE_n \geq \\ & \frac{(1-c)^2}{c} \sum_{k=1}^{\hat{\rho}} \frac{\gamma_k^2 \bar{\gamma}_k}{(\gamma_k + \sigma^2)^2 (\gamma_k + c\sigma^2)^2} + \frac{1}{\sigma^2} \sum_{k=1}^{\hat{\rho}} \frac{\gamma_k \bar{\gamma}_k}{(\gamma_k + \sigma^2)(\gamma_k + c\sigma^2)} \\ & \quad - 2 \frac{1}{\sigma^2} \sum_{k=1}^{\hat{\rho}} \frac{\gamma_k \bar{\gamma}_k}{(\gamma_k + c\sigma^2)^2} - 3c \sum_{k=1}^{\hat{\rho}} \frac{\bar{\gamma}_k}{(\gamma_k + c\sigma^2)^2} \end{aligned} \quad (6.18)$$

almost surely, using the results of Lemma 1 and Baik and Silverstein [2006], Theorem 1.1. Now notice that

$$\begin{aligned} & n \left(\frac{n-p-1}{np} \sum_{c=1}^p \frac{\sigma^2}{l_c} - 1 \right) \\ & = n \left[\frac{n-p-1}{np} \sum_{c=1}^p \frac{\sigma^2}{l_c} - \frac{n-p-1}{n(1-\frac{p-\rho}{n})} + \frac{n-p-1}{n(1-\frac{p-\rho}{n})} - 1 \right] \\ & = \left(1 - \frac{p}{n} - \frac{1}{n} \right) n \left[\frac{1}{p} \sum_{c=1}^p \frac{\sigma^2}{l_c} - \frac{1}{1-\frac{p-\rho}{n}} \right] - \frac{n(\rho+1)}{n-p} \end{aligned} \quad (6.19)$$

and

$$\begin{aligned} & n \left(\frac{1}{1-\frac{p}{n}} - \frac{(n-p-1)(n-p-2)}{n^2 p} \sum_{c=1}^p \frac{\sigma^4}{l_c^2} \right) \\ & = n \left(\frac{n}{n-p} - \frac{(n-p-1)(n-p-2)}{n^2 (1-\frac{p-\rho}{n})^3} \right. \\ & \quad \left. + \frac{(n-p-1)(n-p-2)}{n^2 (1-\frac{p-\rho}{n})^3} - \frac{(n-p-1)(n-p-2)}{n^2 p} \sum_{c=1}^p \frac{\sigma^4}{l_c^2} \right) \end{aligned}$$

$$\begin{aligned}
&= \left(1 - \frac{p}{n} - \frac{1}{n}\right) \left(1 - \frac{p}{n} - \frac{2}{n}\right) n \left[\frac{1}{(1 - \frac{p-\rho}{n})^3} - \frac{1}{p} \sum_{c=1}^p \frac{\sigma^4}{l_c^2} \right] \\
(6.20) \quad &+ \frac{\rho n^2}{(n-p)(n-p+\rho)} + \frac{(2r+3)n^2}{(n-p+\rho)^2} - \frac{(\rho^2+3r+2)n^2}{(n-p+\rho)^3}.
\end{aligned}$$

Moreover,

$$\begin{aligned}
&n \left(\frac{(n-p-1)p}{n^2} \left(\frac{1}{p} \sum_{c=1}^p \frac{\sigma^2}{l_c} \right)^2 - \frac{\frac{p-\rho}{n}}{1 - \frac{p-\rho}{n}} \right) \\
&= n \left(\frac{(n-p-1)p}{n^2} \left(\frac{1}{p} \sum_{c=1}^p \frac{\sigma^2}{l_c} \right)^2 - \frac{(n-p-1)p}{n^2} \frac{1}{(1 - \frac{p-\rho}{n})^2} \right. \\
&\quad \left. + \frac{(n-p-1)p}{n^2} \frac{1}{(1 - \frac{p-\rho}{n})^2} - \frac{\frac{p-\rho}{n}}{1 - \frac{p-\rho}{n}} \right) \\
&= \left(1 - \frac{p}{n} - \frac{1}{n}\right) \frac{p}{n} n \left[\left(\frac{1}{p} \sum_{c=1}^p \frac{\sigma^2}{l_c} \right)^2 - \frac{1}{(1 - \frac{p-\rho}{n})^2} \right] \\
&\quad + \frac{\rho n}{n-p+\rho} - \frac{(\rho+1)pn}{(n-p+\rho)^2} \\
&= \left(1 - \frac{p}{n} - \frac{1}{n}\right) \frac{p}{n} \left[\frac{1}{p} \sum_{c=1}^p \frac{\sigma^2}{l_c} + \frac{1}{1 - \frac{p-\rho}{n}} \right] n \left[\frac{1}{p} \sum_{c=1}^p \frac{\sigma^2}{l_c} - \frac{1}{1 - \frac{p-\rho}{n}} \right] \\
(6.21) \quad &+ \frac{\rho n}{n-p+\rho} - \frac{(\rho+1)pn}{(n-p+\rho)^2}.
\end{aligned}$$

Now, divide the sample covariance matrix as

$$S_n = \sigma^2 \begin{pmatrix} S_n^{11} & S_n^{12} \\ S_n^{21} & S_n^{22} \end{pmatrix}$$

with S_n^{11} $\rho \times \rho$. Then S_n^{22} has a $W_{p-\rho}(n, I)$ distribution – let $\mu_1 > \dots > \mu_{p-\rho}$ be its eigenvalues and notice that by Cauchy's interlacing theorem, $l_i > \sigma^2 \mu_i > l_{i+\rho}$ for all $i = 1, \dots, p - \rho$. Therefore, we have

$$\begin{aligned}
&\frac{p-\rho}{p} n \left[\frac{1}{p-\rho} \sum_{c=1}^{p-\rho} \frac{1}{\mu_c} - \frac{1}{1 - \frac{p-\rho}{n}} \right] + \rho \frac{n}{p} \frac{\sigma^2}{l_p} - \frac{\rho n^2}{(n-p+\rho)p} \\
&\geq n \left[\frac{1}{p} \sum_{c=1}^p \frac{\sigma^2}{l_c} - \frac{1}{1 - \frac{p-\rho}{n}} \right] \geq \\
(6.22) \quad &\frac{p-\rho}{p} n \left[\frac{1}{p-\rho} \sum_{c=1}^{p-\rho} \frac{1}{\mu_c} - \frac{1}{1 - \frac{p-\rho}{n}} \right] + \frac{n}{p} \sum_{c=1}^p \frac{\sigma^2}{l_c} - \frac{\rho n^2}{(n-p+\rho)p}
\end{aligned}$$

and

$$\begin{aligned}
& \frac{p-\rho}{p} n \left[\frac{1}{p-\rho} \sum_{c=1}^{p-\rho} \frac{1}{\mu_c^2} - \frac{1}{(1-\frac{p-\rho}{n})^3} \right] + \rho \frac{n}{p} \frac{\sigma^4}{l_p^2} - \frac{\rho n^4}{(n-p+\rho)^3 p} \\
& \geq n \left[\frac{1}{p} \sum_{c=1}^p \frac{\sigma^4}{l_c^2} - \frac{1}{(1-\frac{p-\rho}{n})^3} \right] \geq \\
(6.23) \quad & \frac{p-\rho}{p} n \left[\frac{1}{p-\rho} \sum_{c=1}^{p-\rho} \frac{1}{\mu_c^2} - \frac{1}{(1-\frac{p-\rho}{n})^3} \right] + \frac{n}{p} \sum_{c=1}^{\rho} \frac{\sigma^4}{l_c^2} - \frac{\rho n^4}{(n-p+\rho)^3 p}.
\end{aligned}$$

Consequently, let us study the quantities

$$n \left[\frac{1}{p-\rho} \sum_{c=1}^{p-\rho} \frac{1}{\mu_c} - \frac{1}{1-\frac{p-\rho}{n}} \right] \quad \text{and} \quad n \left[\frac{1}{p-\rho} \sum_{c=1}^{p-\rho} \frac{1}{\mu_c^2} - \frac{1}{(1-\frac{p-\rho}{n})^3} \right].$$

We use Theorem 1.1 in [Bai and Silverstein \[2004\]](#). First notice that, in our white Wishart case, what they write $F^{c,H}$ is the c.d.f. of a Marčenko-Pastur distribution with parameter c . Let $c_n = (p-\rho)/n$ - then

$$\begin{aligned}
\int \frac{1}{x} dG_n(x) &= n \left[\int \frac{1}{x} dF^{S_n^{22}}(x) - \int \frac{1}{x} dF^{c_n, F^I}(x) \right] \\
&= n \left[\frac{1}{p-\rho} \sum_{c=1}^{p-\rho} \frac{1}{\mu_c} - \int_{[1-\sqrt{c_n}]^2}^{[1+\sqrt{c_n}]^2} \frac{\sqrt{([1+\sqrt{c_n}]^2 - t)(t - [1-\sqrt{c_n}]^2)}}{2\pi c_n t^2} dt \right] \\
&= n \left[\frac{1}{p-\rho} \sum_{c=1}^{p-\rho} \frac{1}{\mu_c} - \frac{1}{(1-\frac{p-\rho}{n})^2} \right]
\end{aligned}$$

and

$$\begin{aligned}
\int \frac{1}{x^2} dG_n(x) &= n \left[\int \frac{1}{x^2} dF^{S_n^{22}}(x) - \int \frac{1}{x^2} dF^{c_n, F^I}(x) \right] \\
&= n \left[\frac{1}{p-\rho} \sum_{c=1}^{p-\rho} \frac{1}{\mu_c^2} - \int_{[1-\sqrt{c_n}]^2}^{[1+\sqrt{c_n}]^2} \frac{\sqrt{([1+\sqrt{c_n}]^2 - t)(t - [1-\sqrt{c_n}]^2)}}{2\pi c_n t^3} dt \right] \\
&= n \left[\frac{1}{p-\rho} \sum_{c=1}^{p-\rho} \frac{1}{\mu_c^2} - \frac{1}{(1-\frac{p-\rho}{n})^3} \right].
\end{aligned}$$

But according to the theorem, as $p_n/n \rightarrow c \in (0, 1)$

$$\left(\int \frac{1}{x} dG_n(x), \int \frac{1}{x^2} dG_n(x) \right) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_2(\tilde{\mu}, \tilde{\Sigma})$$

where the components of $\tilde{\mu}$ and $\tilde{\Sigma}$ are given by eq. (1.6) and (1.7) in the theorem statement ([Bai and Silverstein \[2004\]](#) p. 558). To compute these,

we follow the arguments of Section 5 from the same paper. According to eq. (5.13) from p. 598, we find

$$\begin{aligned}\tilde{\mu}_1 &= \frac{1}{4} \left[\frac{1}{[1 - \sqrt{c}]^2} + \frac{1}{[1 + \sqrt{c}]^2} \right] - \frac{1}{2\pi} \int_{[1 - \sqrt{c}]^2}^{[1 + \sqrt{c}]^2} \frac{dt}{t\sqrt{4c - (t - 1 - c)^2}}, \\ \tilde{\mu}_2 &= \frac{1}{4} \left[\frac{1}{[1 - \sqrt{c}]^4} + \frac{1}{[1 + \sqrt{c}]^4} \right] - \frac{1}{2\pi} \int_{[1 - \sqrt{c}]^2}^{[1 + \sqrt{c}]^2} \frac{dt}{t^2\sqrt{4c - (t - 1 - c)^2}},\end{aligned}$$

and the integrals give, after a Poisson substitution $t = 1 + c - 2\sqrt{c}\cos\theta$,

$$\begin{aligned}\frac{1}{2\pi} \int_{[1 - \sqrt{c}]^2}^{[1 + \sqrt{c}]^2} \frac{dt}{t\sqrt{4c - (t - 1 - c)^2}} &= \frac{1}{2\pi} \int_0^\pi \frac{d\theta}{1 + c - 2\sqrt{c}\cos\theta} \\ &= \frac{1}{2\pi} \left[\frac{2}{1 - c} \arctan \left(\frac{1 + \sqrt{c}}{1 - \sqrt{c}} \tan(\theta/2) \right) \right]_0^\pi \\ &= \frac{1}{2\pi} \left[\frac{2}{1 - c} \frac{\pi}{2} - 0 \right] = \frac{1}{2(1 - c)}, \\ \frac{1}{2\pi} \int_{[1 - \sqrt{c}]^2}^{[1 + \sqrt{c}]^2} \frac{dt}{t^2\sqrt{4c - (t - 1 - c)^2}} &= \frac{1}{2\pi} \int_0^\pi \frac{d\theta}{(1 + c - 2\sqrt{c}\cos\theta)^2} \\ &= \frac{1}{2\pi} \left[\frac{2(1 + c)}{(1 - c)^3} \arctan \left(\frac{1 + \sqrt{c}}{1 - \sqrt{c}} \tan(\theta/2) \right) \right. \\ &\quad \left. + \frac{1}{(1 - c)^2} \frac{2\sqrt{c}\sin\theta}{1 + c - 2\sqrt{c}\cos\theta} \right]_0^\pi \\ &= \frac{1}{2\pi} \left[\frac{2(1 + c)}{(1 - c)^3} \frac{\pi}{2} + 0 - 0 - 0 \right] = \frac{1 + c}{2(1 - c)^3}.\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}\tilde{\mu}_1 &= \frac{1 + c}{2[1 - \sqrt{c}]^2[1 + \sqrt{c}]^2} - \frac{1}{2(1 - c)} = \frac{c}{(1 - c)^2}, \\ \tilde{\mu}_2 &= \frac{1 + 6c + c^2}{2[1 - \sqrt{c}]^4[1 + \sqrt{c}]^4} - \frac{1 + c}{2(1 - c)^3} = \frac{c(c + 3)}{(1 - c)^4}.\end{aligned}$$

For the variances, according to (1.16), p. 564, we can write

$$\begin{aligned}\tilde{\Sigma}_{11} &= -\frac{1}{2\pi^2} \oint_{C_1} \oint_{C_2} \frac{dm_1 dm_2}{z(m_1)z(m_2)(m_1 - m_2)^2}, \\ \tilde{\Sigma}_{12} &= -\frac{1}{2\pi^2} \oint_{C_1} \oint_{C_2} \frac{dm_1 dm_2}{z(m_1)^2 z(m_2)(m_1 - m_2)^2}, \\ \tilde{\Sigma}_{22} &= -\frac{1}{2\pi^2} \oint_{C_1} \oint_{C_2} \frac{dm_1 dm_2}{z(m_1)^2 z(m_2)^2 (m_1 - m_2)^2},\end{aligned}$$

where C_1, C_2 are contours that can be chosen counterclockwise, nonintersecting and enclosing $1/(c - 1)$ (cf. p. 598), and where $z(m)$ stands for the

inverse Stieltjes transform of the complimentary Marčenko-Pastur distribution, which has closed form

$$z(m) = -\frac{1}{m} + \frac{c}{1+m}.$$

We first find:

$$\begin{aligned} \oint_{C_1} \frac{dm_1}{z(m_1)(m_1 - m_2)^2} &= -\frac{1}{1-c} \oint_{C_1} \frac{m_1(m_1 + 1)}{(m_1 - m_2)^2} \frac{dm_1}{m_1 - 1/(c-1)} \\ &= -\frac{2\pi ci}{(1-c)^3} \frac{1}{[m_2 - 1/(c-1)]^2}, \\ \oint_{C_1} \frac{dm_1}{z(m_1)^2(m_1 - m_2)^2} &= \frac{1}{(1-c)^2} \oint_{C_1} \frac{m_1^2(m_1 + 1)^2}{(m_1 - m_2)^2} \frac{dm_1}{[m_1 - 1/(c-1)]^2} \\ &= -4\pi i \frac{c}{(1-c)^5} \frac{1}{[m_2 - 1/(c-1)]^2}. \end{aligned}$$

But then,

$$\begin{aligned} \oint_{C_2} \frac{dm_2}{z(m_2)[m_2 - 1/(c-1)]^2} &= -\frac{1}{1-c} \oint_{C_2} \frac{m_2(m_2 + 1)dm_2}{[m_2 - 1/(c-1)]^3} \\ &= -2\pi i \frac{2}{2!(1-c)} = -2\pi i \frac{1}{1-c} \\ \oint_{C_2} \frac{dm_2}{z(m_2)^2[m_2 - 1/(c-1)]^2} &= \frac{1}{(1-c)^2} \oint_{C_2} \frac{m_2^2(m_2 + 1)^2dm_2}{[m_2 - 1/(c-1)]^4} \\ &= 2\pi i \frac{12(1 - 2/(1-c))}{3!(1-c)^2} = -4\pi i \frac{1+c}{(1-c)^3}. \end{aligned}$$

Therefore,

$$\tilde{\Sigma}_{11} = \frac{2c}{(1-c)^4}, \quad \tilde{\Sigma}_{12} = \frac{4c}{(1-c)^6} \quad \text{and} \quad \tilde{\Sigma}_{22} = \frac{8c(1+c)}{(1-c)^8}.$$

In summary,

$$(6.24) \quad \left(\int \frac{1}{x} dG_n(x), \int \frac{1}{x^2} dG_n(x) \right) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_2 \left(\begin{bmatrix} \frac{c}{(1-c)^2} \\ \frac{c(c+3)}{(1-c)^4} \end{bmatrix}, \begin{bmatrix} \frac{2c}{(1-c)^4} & \frac{4c}{(1-c)^6} \\ \frac{4c}{(1-c)^6} & \frac{8c(1+c)}{(1-c)^8} \end{bmatrix} \right).$$

Therefore, going back to eq. (6.17) and ineq. (6.19), (6.20), (6.21), (6.22) and (6.23), we have the upper bound

$$\begin{aligned}
& n(\tilde{\sigma}^2 - \sigma^2) \\
&= \frac{\sigma^2}{\sigma^4 B} \left[\left(1 - \frac{p}{n} - \frac{1}{n}\right) n \left[\frac{1}{p} \sum_{c=1}^p \frac{\sigma^2}{l_c} - \frac{1}{1 - \frac{p-\rho}{n}} \right] \right. \\
&\quad \left. - \frac{n(\rho+1)}{n-p} \right. \\
&\quad \left. - \left(1 - \frac{p}{n} - \frac{1}{n}\right) \left(1 - \frac{p}{n} - \frac{2}{n}\right) n \left[\frac{1}{p} \sum_{c=1}^p \frac{\sigma^4}{l_c^2} - \frac{1}{(1 - \frac{p-\rho}{n})^3} \right] \right. \\
&\quad \left. + \frac{\rho n^2}{(n-p)(n-p+\rho)} + \frac{(2\rho+3)n^2}{(n-p+\rho)^2} - \frac{(\rho^2+3\rho+2)n^2}{(n-p+\rho)^3} \right. \\
&\quad \left. + \left(1 - \frac{p}{n} - \frac{1}{n}\right) \frac{p}{n} \left[\frac{1}{p} \sum_{c=1}^p \frac{\sigma^2}{l_c} + \frac{1}{1 - \frac{p-\rho}{n}} \right] n \left[\frac{1}{p} \sum_{c=1}^p \frac{\sigma^2}{l_c} - \frac{1}{1 - \frac{p-\rho}{n}} \right] \right. \\
&\quad \left. + \frac{rn}{n-p+\rho} - \frac{(\rho+1)pn}{(n-p+\rho)^2} \right] + \frac{nE_n}{B} \\
&\leq \frac{\sigma^2}{\sigma^4 B} \left[\left(1 - \frac{p}{n} - \frac{1}{n}\right) \frac{p-\rho}{p} n \left[\frac{1}{p-\rho} \sum_{c=1}^{p-\rho} \frac{1}{\mu_c} - \frac{1}{1 - \frac{p-\rho}{n}} \right] - \frac{n(\rho+1)}{n-p} \right. \\
&\quad \left. + \left(1 - \frac{p}{n} - \frac{1}{n}\right) \rho \frac{n}{p} \frac{\sigma^2}{l_p} - \left(1 - \frac{p}{n} - \frac{1}{n}\right) \frac{\rho n^2}{(n-p+\rho)p} \right. \\
&\quad \left. - \left(1 - \frac{p}{n} - \frac{1}{n}\right) \left(1 - \frac{p}{n} - \frac{2}{n}\right) \frac{p-\rho}{p} n \left[\frac{1}{p-\rho} \sum_{c=1}^{p-\rho} \frac{1}{\mu_c^2} - \frac{1}{(1 - \frac{p-\rho}{n})^3} \right] \right. \\
&\quad \left. - \left(1 - \frac{p}{n} - \frac{1}{n}\right) \left(1 - \frac{p}{n} - \frac{2}{n}\right) \frac{n}{p} \sum_{c=1}^{\rho} \frac{\sigma^4}{l_c^2} \right. \\
&\quad \left. + \left(1 - \frac{p}{n} - \frac{1}{n}\right) \left(1 - \frac{p}{n} - \frac{2}{n}\right) \frac{\rho n^4}{(n-p+\rho)^3 p} \right. \\
&\quad \left. + \frac{\rho n^2}{(n-p)(n-p+\rho)} + \frac{(2\rho+3)n^2}{(n-p+\rho)^2} - \frac{(\rho^2+3\rho+2)n^2}{(n-p+\rho)^3} \right. \\
&\quad \left. + \left(1 - \frac{p}{n} - \frac{1}{n}\right) \frac{p}{n} \left[\frac{1}{p} \sum_{c=1}^p \frac{\sigma^2}{l_c} + \frac{1}{1 - \frac{p-\rho}{n}} \right] \right. \\
&\quad \left. \frac{p-\rho}{p} n \left[\frac{1}{p-\rho} \sum_{c=1}^{p-\rho} \frac{1}{\mu_c} - \frac{1}{1 - \frac{p-\rho}{n}} \right] \right. \\
&\quad \left. + \left(1 - \frac{p}{n} - \frac{1}{n}\right) \frac{p}{n} \left[\frac{1}{p} \sum_{c=1}^p \frac{\sigma^2}{l_c} + \frac{1}{1 - \frac{p-\rho}{n}} \right] \rho \frac{n}{p} \frac{\sigma^2}{l_p} \right]
\end{aligned}$$

$$\begin{aligned}
& - \left(1 - \frac{p}{n} - \frac{1}{n}\right) \frac{p}{n} \left[\frac{1}{p} \sum_{c=1}^p \frac{\sigma^2}{l_c} + \frac{1}{1 - \frac{p-\rho}{n}} \right] \frac{\rho n^2}{(n-p+\rho)p} \\
& + \frac{\rho n}{n-p+\rho} - \frac{(\rho+1)pn}{(n-p+\rho)^2} \Big] + \frac{nE_n}{B} \\
& = \frac{\sigma^2}{\sigma^4 B} \left(a_1^{(n)} n \left[\frac{1}{p-\rho} \sum_{c=1}^{p-\rho} \frac{1}{\mu_c} - \frac{1}{1-c} \right] + a_2^{(n)} n \left[\frac{1}{p-\rho} \sum_{c=1}^{p-\rho} \frac{1}{\mu_c^2} - \frac{1}{(1-c)^3} \right] \right. \\
& \quad \left. + b^{(n)} \right) + \frac{nE_n}{B}.
\end{aligned}$$

Now note that

$$\begin{aligned}
& a_1^{(n)} \xrightarrow{\text{a.s.}} 1+c, \quad a_2^{(n)} \xrightarrow{\text{a.s.}} -(1-c)^2, \\
& b^{(n)} \xrightarrow{\text{a.s.}} \frac{2c(\rho+1)-1}{(1-c)^2} + \frac{(2-c)(1+c)\rho}{(1-\sqrt{c})^2} - \frac{(1-c)^2}{c} \sum_{k=1}^{\rho} \frac{\sigma^4 \gamma_k^2}{(\gamma_k + \sigma^2)^2 (\gamma_k + c\sigma^2)^2}
\end{aligned}$$

using Lemma 1 and Baik and Silverstein [2006], Theorem 1.1. Therefore, using Slutsky and eq. (6.24),

$$\begin{aligned}
& a_1^{(n)} n \left[\frac{1}{p-\rho} \sum_{c=1}^{p-\rho} \frac{1}{\mu_c} - \frac{1}{1-c} \right] + a_2^{(n)} n \left[\frac{1}{p-\rho} \sum_{c=1}^{p-\rho} \frac{1}{\mu_c^2} - \frac{1}{(1-c)^3} \right] + b^{(n)} \\
& \xrightarrow{\mathcal{D}} \text{N} \left(\mu, \frac{2c(1+c)^2}{(1-c)^4} \right),
\end{aligned}$$

with

$$\mu = \frac{(2-c)(1+c)\rho}{(1-\sqrt{c})^2} + \frac{2c\rho-1}{(1-c)^2} - \frac{(1-c)^2}{c} \sum_{k=1}^{\rho} \frac{\sigma^4 \gamma_k^2}{(\gamma_k + \sigma^2)^2 (\gamma_k + c\sigma^2)^2}.$$

Therefore, using eq. (6.16) and (6.18) we obtain $n(\tilde{\sigma}^2 - \sigma^2) \leq X_n^+$ with $X_n^+ \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \text{N} \left(\mu^+, \frac{2c(1+c)^2 \sigma^4}{(1-c)^4} \right)$, where

$$\begin{aligned}
\mu^+ & = \frac{(2c\rho-1)\sigma^2}{(1-c)^2} + \frac{(2-c)(1+c)\rho\sigma^2}{(1-\sqrt{c})^2} - \frac{(1-c)^2}{c} \sum_{k=1}^{\rho} \frac{\sigma^6 \gamma_k^2}{(\gamma_k + \sigma^2)^2 (\gamma_k + c\sigma^2)^2} \\
& \quad + \frac{(1-c)^2}{c} \sum_{k=1}^{\rho} \frac{\gamma_k^2 \tilde{\gamma}_k \sigma^4}{(\gamma_k + \sigma^2)^2 (\gamma_k + c\sigma^2)^2} + \sum_{k=1}^{\rho} \frac{\gamma_k \tilde{\gamma}_k \sigma^2}{(\gamma_k + \sigma^2)(\gamma_k + c\sigma^2)} \\
(6.25) \quad & - 2 \sum_{k=1}^{\rho} \frac{\gamma_k \tilde{\gamma}_k \sigma^2}{(\gamma_k + c\sigma^2)^2}.
\end{aligned}$$

The same argument can be done to obtain a lower bound. Using eq. (6.17), (6.19), (6.20), (6.21), (6.22) and (6.23) again, we get

$$\begin{aligned} n(\tilde{\sigma}^2 - \sigma^2) &\leq \frac{\sigma^2}{\sigma^4 B} \left(a_1^{(n)} n \left[\frac{1}{p-\rho} \sum_{c=1}^{p-\rho} \frac{1}{\mu_c} - \frac{1}{1-c} \right] \right. \\ &\quad \left. + a_2^{(n)} n \left[\frac{1}{p-\rho} \sum_{c=1}^{p-\rho} \frac{1}{\mu_c^2} - \frac{1}{(1-c)^3} \right] + b^{(n)} \right) + \frac{nE_n}{B} \end{aligned}$$

where

$$\begin{aligned} a_1^{(n)} &\xrightarrow{\text{a.s.}} 1+c, \quad a_2^{(n)} \xrightarrow{\text{a.s.}} -(1-c)^2, \\ b^{(n)} &\xrightarrow{\text{a.s.}} \frac{2c(\rho+1)-1}{(1-c)^2} - \frac{(1-c)^2\rho}{c(1-\sqrt{c})^2} + \frac{1+c}{c} \sum_{k=1}^{\rho} \frac{\sigma^2\gamma_k}{(\gamma_k+\sigma^2)(\gamma_k+c\sigma^2)} \end{aligned}$$

Therefore, again using eq. (6.16) and (6.18) we obtain that $n(\tilde{\sigma}^2 - \sigma^2) \geq X_n^-$ with $X_n^- \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N\left(\mu^-, \frac{2c(1+c)^2\sigma^4}{(1-c)^4}\right)$, where

$$\begin{aligned} \mu^- &= \frac{(2c\rho-1)\sigma^2}{(1-c)^2} - \frac{(1-c)^2\rho\sigma^2}{c(1-\sqrt{c})^2} + \frac{1+c}{c} \sum_{k=1}^{\rho} \frac{\sigma^4\gamma_k}{(\gamma_k+\sigma^2)(\gamma_k+c\sigma^2)} \\ &\quad + \frac{(1-c)^2}{c} \sum_{k=1}^{\rho} \frac{\gamma_k^2\bar{\gamma}_k\sigma^4}{(\gamma_k+\sigma^2)^2(\gamma_k+c\sigma^2)^2} + \sum_{k=1}^{\rho} \frac{\gamma_k\bar{\gamma}_k\sigma^2}{(\gamma_k+\sigma^2)(\gamma_k+c\sigma^2)} \\ (6.26) \quad &\quad - 2 \sum_{k=1}^{\rho} \frac{\gamma_k\bar{\gamma}_k\sigma^2}{(\gamma_k+c\sigma^2)^2} - 3c \sum_{k=1}^{\rho} \frac{\bar{\gamma}_k\sigma^4}{(\gamma_k+c\sigma^2)^2}. \end{aligned}$$

This concludes the proof. \square

Proof of Lemma 2. Let $\Sigma' \in B_r(\Sigma, 2M)$, and write $\lambda_i = \lambda_i(\Sigma)$, $\lambda'_i = \lambda'_i(\Sigma'_p)$ to simplify notation. Since the sequences are spiked, there are a finite number of different eigenvalues as $n \rightarrow \infty$. We can decompose

$$N(0, \Sigma_p)^n = \bigotimes_{i=1}^p N(0, \lambda_i)^n, \quad N(0, \Sigma'_p)^n = \bigotimes_{i=1}^p N(0, \lambda'_i)^n.$$

Let λ be the Lebesgue measure on \mathbb{R} . Writing the Hellinger affinity between two densities f, g as $\alpha(f, g) = \int \sqrt{fg} d\lambda$, we have by Cauchy-Schwarz

$$\begin{aligned} \delta_{TV}(f, g)^2 &= 4 \left(\int |f - g| d\lambda \right)^2 \leq \frac{1}{4} (1 - \alpha(f, g)) (1 + \alpha(f, g)) \\ &= (1 - \alpha(f, g))^2 \end{aligned}$$

Now for two normals, we have

$$\alpha(N(0, a^2), N(0, b^2))^2 = \frac{2ab}{a^2 + b^2}$$

and since Hellinger affinity distributes over a product of independent densities, we get

$$\delta_{TV} \left(N(0, \Sigma_p)^n, N(0, \Sigma'_p)^n \right)^2 \leq 1 - \prod_{i=1}^p \left(\frac{2\sqrt{\lambda_i \lambda'_i}}{\lambda_i + \lambda'_i} \right)^n$$

Now note that $\frac{2\sqrt{\lambda_i \lambda'_i}}{\lambda_i + \lambda'_i} > \frac{\sqrt{1-2M/n^r}}{1-M/n^r}$ if and only if $\lambda_i(1 - \frac{2M}{n^r}) < x < \lambda_i(1 + \frac{2M}{n^r(1-2M/n^r)})$. By definition, $\Sigma' \in B_r(\Sigma, 2M)$ means $|\lambda_i - \lambda'_i| < \frac{2M\lambda_i}{n^r}$ for all i , so we have the bound

$$< 1 - \left(\frac{\sqrt{1-2M/n^r}}{1-M/n^r} \right)^{np} = 1 - \left(1 - \frac{M^2}{n^{2r}(1-M/n^r)^2} \right)^{np/2}$$

over all $\Sigma' \in B_r(\Sigma, 2M)$. Taking a supremum and a limit yields

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sup_{\Sigma' \in B_1(\Sigma, 2M)} \delta_{TV} \left(N(0, \Sigma_p)^n, N(0, \Sigma'_p)^n \right)^2 \\ & \leq \lim_{n \rightarrow \infty} \left[1 - \left(1 - \frac{M^2}{n^2(1-M/n)^2} \right)^{np/2} \right] = 1 - e^{-cM^2/2} \end{aligned}$$

and

$$\lim_{n \rightarrow \infty} \sup_{\Sigma' \in B_r(\Sigma, 2M)} \delta_{TV} \left(N(0, \Sigma_p)^n, N(0, \Sigma'_p)^n \right)^2 = 0,$$

as desired. \square

Proof of Theorem 3. Define $\Sigma'_p = (1 - \frac{2M_\epsilon}{n})\Sigma_p$. We have

$$(6.27) \quad |\sigma^2 - \sigma^{2'}| = 2\sigma^2 \frac{M_\epsilon}{n}.$$

Moreover, according to Lemma 2, we have

$$\lim_{n \rightarrow \infty} \delta_{TV} \left(N(0, \Sigma_p)^n, N(0, \Sigma'_{p,M})^n \right) \leq \sqrt{1 - \exp\left(-\frac{cM_\epsilon^2}{2}\right)} = 1 - 4\epsilon,$$

so for some N_ϵ we have $\delta_{TV} < 1 - 2\epsilon$ for all $n \geq N_\epsilon$. Now note that $\Sigma'_p \in B_1(\Sigma_p, \frac{2M_\epsilon}{n}\|\Sigma_p\|_2)$. Say for some $\hat{\sigma}^2$ we have

$$(6.28) \quad \sup_{\Sigma' \in B_1(\Sigma, 2M_\epsilon)} P_{\Sigma'_p} \left[|\hat{\sigma}^2 - \sigma^{2'}| \geq \sigma^2 \frac{M_\epsilon}{n} \right] < \epsilon.$$

Define the event $A = [|\hat{\sigma}^2 - \sigma^2| \leq \sigma^2 \frac{M_\epsilon}{n}]$ - then by (6.27) and (6.28) we find $P_{\Sigma_p}[A] \geq 1 - \epsilon$ and $P_{\Sigma'_p}[A] < \epsilon$. Therefore,

$$\delta_{TV} \left(N(0, \Sigma_p)^n, N(0, \Sigma'_{p,M})^n \right) \geq P_{\Sigma_p}[A] - P_{\Sigma'_p}[A] \geq 1 - 2\epsilon$$

which contradicts $\delta_{TV} < 1 - 2\epsilon$ for all $n \geq N_\epsilon$ - therefore,

$$\sup_{\Sigma' \in B_1(\Sigma, 2M_\epsilon)} \mathbb{P}_{\Sigma'_p} \left[|\hat{\sigma}^2 - \sigma^{2'}| \geq \sigma^2 \frac{M_\epsilon}{n^r} \right] \geq \epsilon$$

for all $\hat{\sigma}^2$ and $n \geq N_\epsilon$, yielding the desired result. \square

Proof of Proposition 2. We have:

$$\begin{aligned} \mathbb{P}_{\Sigma'_p} \left[|\tilde{\sigma}^2 - \sigma^{2'}| \geq \sigma^2 \frac{M}{n^r} \right] &\leq \mathbb{P}_{\Sigma_p} \left[|\tilde{\sigma}^2 - \sigma^{2'}| \geq \sigma^2 \frac{M}{n^r} \right] \\ &+ \left| \mathbb{P}_{\Sigma_p} \left[|\tilde{\sigma}^2 - \sigma^{2'}| \geq \sigma^2 \frac{M}{n^r} \right] - \mathbb{P}_{\Sigma'_p} \left[|\tilde{\sigma}^2 - \sigma^{2'}| \geq \sigma^2 \frac{M}{n^r} \right] \right| \\ &= A_p + B_p. \end{aligned}$$

Choose any $\Sigma' \in B_r(\Sigma, 2M)$ and write $\lambda_i = \lambda_i(\Sigma)$, $\lambda'_i = \lambda'_i(\Sigma')$. Using Lemma 2, we find for the second term

$$\begin{aligned} \lim_{n \rightarrow \infty} \sup_{\Sigma \in B_r(\Sigma, 2M)} B_p &\leq \lim_{n \rightarrow \infty} \sup_{\Sigma' \in B_r(\Sigma, 2M)} \delta_{TV} \left(N(0, \Sigma_p)^n, N(0, \Sigma'_p)^n \right) \\ &= 0. \end{aligned}$$

For the first term, let us use Theorem 2. Since $|\sigma^{2'} - \sigma^2| < \sigma^2 M/n^r$, we have

$$\begin{aligned} \sup_{\Sigma \in B(\Sigma, 2M)} A_p &\leq \mathbb{P}_{\Sigma_p} \left[n |\tilde{\sigma}^2 - \sigma^2| > \sigma^2 \frac{M}{n^{r-1}} \right] \\ &\leq \mathbb{P}_{\Sigma_p} \left[\frac{(1-c)^2}{\sqrt{2c}(1+c)\sigma^2} (X_n^+ - \mu^+) > \frac{(1-c)^2(-2\sigma^2\mu^+ + M/n^{r-1})}{2\sqrt{2c}(1+c)} \right] \\ &\quad + \mathbb{P}_{\Sigma_p} \left[\frac{(1-c)^2}{\sqrt{2c}(1+c)\sigma^2} (X_n^- - \mu^-) < \frac{(1-c)^2(-2\sigma^2\mu^- - M/n^{r-1})}{2\sqrt{2c}(1+c)} \right] \\ &\xrightarrow{n \rightarrow \infty} 1 - \Phi(\infty) + \Phi(-\infty) = 0. \end{aligned}$$

This concludes the proof. \square

6.3. Proofs for Section 4.

Proof of proposition 3. We first notice that, from Lemma 1 and Theorem 2, for any $1 \leq k \leq \rho$ we have

$$(6.29) \quad \hat{\gamma}_{\rho,k} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \gamma_k, \quad \tilde{\sigma}_\rho^2 \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \sigma^2.$$

Denote as usual the unbiased risk estimator of Theorem 1 at $\hat{\Gamma}_r + \tilde{\sigma}_r^2 I_p$ by $F_r + G_r$. By their definitions and Lemma 1, we see that

$$F_\rho + G_\rho \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 0.$$

Thus there exists a N_1 (random) such that $\rho \in \left\{r \mid |F_r + G_r| \leq \frac{p+1}{n}\right\}$ for all $n \geq N_1$. Now note that there is a N_2 (also random) such that for all $n \geq N_2$, $\left\{\frac{\mathbb{1}_{[r \leq p]}}{l_{r+1}} \frac{(1+\sqrt{p/n})^2}{p-r} \sum_{c=r+1}^p l_c \geq 1\right\} = \{\rho\}$. Thus for all $n \geq N_1 \vee N_2$,

$$\hat{\rho} = \min\{\rho\} = \rho,$$

so in particular $\hat{\rho} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \rho$. Thus, by eq. (6.29) and any $\epsilon > 0$ there exists a N_3 random such that for all $n \geq N_1 \vee N_2 \vee N_3$, $k < p$ and $|\hat{\gamma}_{\hat{\rho},k} - \gamma_k| = |\hat{\gamma}_{\rho,k} - \gamma_k| < \epsilon$. That is, $\gamma_{\hat{\rho},k} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \gamma_k$. Finally, again by eq. (6.29) for all $\epsilon > 0$, there exists a N_3 random such that for all $n \geq N_1 \vee N_2 \vee N_3$, $|\tilde{\sigma}^2 - \sigma^2| = |\tilde{\sigma}_{\hat{\rho}}^2 - \sigma^2| = |\tilde{\sigma}_{\rho}^2 - \sigma^2| < \epsilon$, i.e. $\tilde{\sigma}^2 \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \sigma^2$, as desired. \square

REFERENCES

- Z.D. Bai and J.W. Silverstein. Clt for linear spectral statistics of large-dimensional sample covariance matrices. *Annals of Probability*, 32:553–605, 2004.
- J. Baik and J.W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97:1382–1408, 2006.
- J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability*, pages 1643–1697, 2005.
- P. J. Bickel and E. Levina. Regularized estimation of large covariance matrices. *Annals of Statistics*, pages 199–227, 2008a.
- P. J. Bickel and E. Levina. Covariance regularization by thresholding. *Annals of Statistics*, pages 2577–2604, 2008b.
- J. Bien, F. Bunea, and L. Xiao. Convex banding of the covariance matrix. *arXiv preprint arXiv:1405.6210*, 2014.
- T. Cai and W. Liu. Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106:672–684, 2011.
- T. Cai, W. Liu, and X. Luo. A constrained l1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106:594–607, 2011.
- D. Chételat and M.T. Wells. Improved multivariate normal mean estimation with unknown covariance when p is greater than n . *Annals of Statistics*, 40:3137–3160, 2012.
- L.D. Donoho, M. Gavish, and I. Johnstone. Optimal shrinkage of eigenvalues in the spiked covariance model. *arXiv preprint arXiv:1311.0851*, 2014.
- N. El Karoui. Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Annals of Statistics*, 36:2757–2790, 2008a.
- N. El Karoui. Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Annals of Statistics*, pages 2717–2756, 2008b.

- J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9:432–441, 2008.
- M. Giaquinta and S. Hildebrandt. *Calculus of Variations*, volume I. Springer, Berlin, 1996.
- L.R. Haff. Minimax estimators for a multinormal precision matrix. *Journal of Multivariate Analysis*, 7:374–385, 1977.
- L.R. Haff. An identity for the wishart distribution with applications. *Journal of Multivariate Analysis*, 9:531–544, 1979.
- L.R. Haff. Empirical bayes estimation of the multivariate normal covariance matrix. *Annals of Statistics*, 8:586–597, 1980.
- L.R. Haff. The variational form of certain bayes estimators. *Annals of Statistics*, 19:1109–1680, 1991.
- W. James and C. Stein. Estimation with quadratic loss. In *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, volume 1, pages 361–380. Berkeley, University of California Press, 1961.
- I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, pages 295–327, 2001.
- I. M. Johnstone and A. Y. Lu. On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association*, 104, 2009.
- Y. Konno. Shrinkage estimators for large covariance matrices in multivariate real and complex normal distributions under an invariant quadratic loss. *Journal of Multivariate Analysis*, 100:2237–2253, 2009.
- T. Kubokawa and M.S. Srivastava. Estimation of the precision matrix of a singular wishart distribution and its application in high-dimensional data. *Journal of Multivariate Analysis*, 99:1906–1928, 2008.
- L. Le Cam. Convergence of estimates under dimensionality restrictions. *Annals of Statistics*, 1:38–53, 1973.
- O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88:365–411, 2004.
- O. Ledoit and M. Wolf. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics*, 40:1024–1060, 2012.
- S.P. Lin and M.D. Perlman. A monte carlo comparison of four estimators of a covariance matrix. In P.R. Krishnaiah, editor, *Multivariate Analysis VI*, pages 411–429. Elsevier Science Publishers B.V., Amsterdam, 1985.
- N. Meinshausen and P. Bühlmann. High-dimensional graphs and variable selection with the lasso. *Annals of Statistics*, pages 1436–1462, 2006.
- R.J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley, New York, 1982.
- B. Nadler. Finite sample approximation results for principal component analysis: a matrix perturbation approach. *Annals of Statistics*, 36:2791–2817, 2008.
- Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, 17(4):1617, 2007.

- P. Ravikumar, M. J. Wainwright, G. Raskutti, and B. Yu. High-dimensional covariance estimation by minimizing 1-penalized log-determinant divergence. *Electronic Journal of Statistics*, 5:935–980, 2011.
- A. J. Rothman, P. J. Bickel, E. Levina, and J. Zhu. Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, 2:494–515, 2008.
- A. J. Rothman, E. Levina, and J. Zhu. Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association*, 104: 177–186, 2009.
- Y. Sheena. Unbiased estimator of risk for an orthogonally invariant estimator of a covariance matrix. *Journal of the Japan Statistical Society*, 25:35–48, 1995.
- C. Stein. Estimation of a covariance matrix. In *Rietz Lecture, 39th Annual Meeting of the IMS*. Atlanta, Georgia, 1975.
- C. Stein. Lectures on the theory of estimation of many parameters. *Journal of Soviet Mathematics*, 34:1373–1403, 1986.